Essays on noninvertible ARMA models

by

Juho Nyholm

MSSc

Academic dissertation to be presented for public examination,
by due permission of the Faculty of Social Sciences
of the University of Helsinki,
in the Lecture Hall of Economicum, Arkadiankatu 7,
on Friday, 15 November 2019 at 12 noon.

Research Reports
Publications of the Helsinki Center of Economic Research, No. 2019:3
Dissertationes Oeconomicae

Juho Nyholm

Essays on noninvertible ARMA models

ISBN: 978-952-10-8752-3 (print)
ISBN: 978-952-10-8753-0 (online)
ISSN: 2323-9786 (print)
ISSN: 2323-9794 (online)

Unigrafia Oy
Helsinki 2019

Abstract

The theory for conventional Gaussian, causal and invertible autoregressive moving average (ARMA) models has developed into a what can be considered as a basis of modern time series analysis. The concept of noninvertibility is plausible only under non-Gaussian processes because non-Gaussianity is a necessary condition for the statistical identification of the noninvertible ARMA model. Therefore, if the Gaussianity assumption is relaxed, we can study a richer class of models which are, unlike their invertible counterparts, capable of capturing nonlinear patterns in the data.

The aim of this thesis is to consider some of the novel results in ARMA modeling of stationary time series data, and to expand these results to a particular case of noninvertibility and non-Gaussianity of the model. It also aims at providing insights on applicability of noninvertible ARMA models in financial time series analysis.

The first essay proposes two residual-based diagnostic tests for noninvertible ARMA models. The tests are analogous to the portmanteau tests developed by Box and Pierce (1970) and McLeod and Li (1983) in the conventional invertible case. We derive the asymptotic χ^2 distribution for the tests under the null of correctly specified model, and study the size and power properties in a Monte Carlo simulation study. An empirical application employing financial time series data points out the usefulness of noninvertible ARMA model in analyzing stock returns and the use of the proposed test statistics.

The second essay studies properties of the maximum likelihood estimator of a noninvertible ARMA model with errors that follow an α-stable distribution and have infinite variance. To ensure the identification of the noninvertible ARMA model considered, we restrict the analysis to non-Gaussian distributions. Estimators of the autoregressive and moving average parameters are shown to be $n^{1/\alpha}$-consistent and to converge to a non-standard limiting distribution that is obtained as a maximizer of a certain random function. Estima-

tors of the parameters in the α-stable distribution have the conventional $n^{1/2}$ rate of convergence. Finite sample properties of the estimators are studied in a simulation experiment, and an application to financial trading volume data illustrates the applicability of the model.

The third and last essay looks for nonlinear predictability in stock returns. For many theoretical asset pricing models, predictability follows as an implication of risk aversion of agents. A closed form solutions for the next periods asset return depends on how the agents form their expectations about the future state of the world. By no means should this predictability be linear. First, we provide evidence of predictability of returns of U.S. stock portfolios and individual financial sector stocks using noninvertible ARMA(1,1) model and two-stage predictability testing procedure. Second, we provide a straightforward extensions to this procedure and allow for a larger model than the non-invertible ARMA(1,1).

Acknowledgments

There are many people I wish to thank for helping me get through my dissertation project. First of all, I want to thank my supervisor, Professor Mika Meitz, for inspiring me with your own research on time series econometrics, and for a patient and encouraging guidance. No matter what kind of a dead end I had reached, after discussing with you, I was always well aware of the next steps I was ought to take. For making me think clearer and concentrate on the things that mattered, you have my gratitude.

I am very thankful to Associate Professor Beth Andrews and Professor Alain Hecq for a careful pre-examination of this work, and for constructive comments that improved the quality of this dissertation substantially. I am also thankful to Associate Professor Rickard Sandberg for agreeing to act as the opponent in my public defense.

Much of this thesis is devoted to the asymptotic theory in time series statistics, and for everything I know about this topic, I am in dept to Professor Emeritus Pentti Saikkonen. Your classes and all the conversations with you laid down the foundation that allowed me to overtake this project. Your role in guiding me towards my research questions was also indispensable. This work would not have bee accomplished without support from Professor Markku Lanne. Not the least because of your dedication to the *Research Group in Financial and Macroeconometrics*, under which I was privileged to carry out my research agenda for several years in a doctoral student position. Your detailed comments and suggestions on the papers and invaluable help with all the practicalities are gratefully acknowledged.

I am much obliged to the Finnish time series econometrics community for many comments and discussion over the past years. Docent Henri Nyberg has helped me substantially by reading and commenting my papers at various stages. I want to thank all the participants in many HECER (Helsinki

GSE) Econometric Workshops over these past years. I have also benefited greatly from the comments by Assistant Professor Matthijs Lof and Professor Niklas Ahlgren. Discussions with Bernd Funovits opened up a whole new perspective on the noninvertible processes. You have my gratitude.

Part of this project was carried out during my visit in the Department of Economics at Duke University. I want to thank Professors Tim Bollerslev and Andrew Patton for your time and hospitality. I was also given a chance to participate in their Financial Econometrics Lunch Group and present my own research there. I am thankful to all the members of that community for your comments.

Financial support from the Yrjö Jahnsson Foundation, the O-P Group Research Foundation, the Research Funds of the University of Helsinki, and the Academy of Finland is gratefully acknowledged, as well as the Chancellor's travel grants. I am grateful for your generosity.

I thank Jaakko Nelimarkka for being such a good example to me from the day one of this endeavor. My years in the same office with you were nothing short of inspirational and enjoyable, for which I am grateful to you. My life as a doctoral student would have not been as pleasant as it was without my friends and colleagues at Economicum. I want to thank each and everyone of you for peer support and intriguing conversations over lunch breaks throughout these years. I would like to thank Tuomas Nurminen and Timo Autio for taking my mind off this project from time to time, and for finding time for our hobbies outside our offices.

Lastly, I want to express my gratitude to my dear Tiina. It is finally the time for us to meet without this thing on my shoulders. Finishing this project was made possible by your caring support and endless understanding.

Helsinki, October 2019

Juho Nyholm

Contents

1 Introduction

1.1 Background

In their seminal book, Box and Jenkins (1970) introduce a time series modeling framework which is nowadays known as Box-Jenkings methodology. Their approach is, first, to consider the class of models which are sufficient in controlling the features of the data. For example, the models need to be able to overcome nonlinearities, seasonalities, or nonstationarity of the data. Second, the class of statistical models must be related to the observed data. One needs to identify the correct form of the model. This can be done, for example, by investigating the autocorrelation and partial autocorrelation structure of the data. Third, the identified model must be fitted to the data. This stage is concerned with estimating the model parameters using some estimation method, maximum likelihood for example. In the fourth and last step, one must conclude if the chosen model indeed provides an adequate fit to the data. For this reason, one may use diagnostic checks to judge, if the residuals of the fitted model contain information that was not used in the modeling of the data, and if so, the model should be modified to incorporate this information.

Autoregressive moving average (ARMA) class of models play a key role in analysis of stationary time series data in various fields. Since Box-Jenkins methodology was introduced, a vast literature on estimation theory, predicting time series data, model selection, and identification has been developed. This thesis expands on some of the novel results in conventional invertible ARMA modeling, and generalizes these results for a particular case of noninvertibility and non-Gaussianity. In particular, these results contribute to the third and fourth steps in Box-Jenkins methodology: We provide means for checking adequacy of an estimated noninvertible ARMA model under non-Gaussian error terms, and we provide maximum likelihood theory for noninvertible ARMA models under the assumption of a heavy tailed error process. Thesis also contains an empirical study, in which the time series properties of financial data is studied using noninvertible ARMA models, illustrating the usefulness of noninvertible models in applied econometric research.

The results of this thesis are motivated by financial market data. There are two point of view as to, why these methods should be of interest to practitioners. The first is derived from economic theory. In recent years, there has been a considerable branch of economic literature aimed at modeling different areas of economy, and their linkages, in which the solution of the models turns out to be noninvertible (nonfundamental). Although the mapping between noninvertible statistical models and nonfundamental solutions of the economic models is ambiguous, there is already some evidence for the applicability for these types of models in situations where we could expect that the data is generated by a noninvertible process (see for example Lof, 2013, for a small scale simulation study).

The second reason why we should be motivated to develop noninvertible models in order to understand financial data better, comes from the properties of the models. There are some features in the financial times series data that are consistently encountered, and some properties that the models should be able to model. First, financial data seems to be predominantly heteroskedastic. This feature is so common that there exists a vast literature on nonlinear time series models that have been developed to capture the the clustering of the volatility. Any statistical model that is aimed at modeling financial data, should be able to incorporate this feature. Second, Gaussianity might not be a good description of the distribution of the financial data; the tails of the distribution seem to be heavier than what one would expect if the process were Gaussian. Although non-Gaussianity is a prerequisite for the identification of the noninvertible model, this assumption does not seem to be very restrictive in our case. To the contrary, in Chapters 2 and 4 we estimate the model using quarterly stock return data with Student's *t*-distribution. Estimation results suggests that the kurtosis of the error distribution is considerable larger than that of the Gaussian distribution. In Chapter 3, we go even further, and show how to estimate this model without making any assumption on the finiteness of the moments, and we assume that the distribution belongs in the attraction of domain of stable distributions.

The third feature of the data we wish to explain, is the predictability. This need arises from the theoretical asset pricing literature (see for example Chapter 9 in Singleton, 2009). Dynamic asset pricing literature suggests, that risk aversion should impose predictability in the asset returns. This predictability, however, might not be easy to capture, as it might be nonlinear as well as linear. Noninvertibel ARMA processes are always predictable. Unlike their conventional invertible counterparts, there is nonlinear predictability present even if the observations are not autocorrelated.

1.2 The noninvertible ARMA model

This section introduces the noninvertible ARMA model that will be present in all of the Chapters of this thesis. The model specification is adapted from Meitz and Saikkonen (2013), which studied a noninvertible model with an ARCH type heteroskedasticity. After the introduction to the model, we will briefly discuss the identification of the model, especially the importance of the assumption of non-Gaussianity. We will then derive a backward looking presentation for this model in order to illustrate the nonlinear nature of this process. We will briefly talk about the predictability of noninvertible ARMA processes and then we illustrate graphically some of these features.

1.2.1 Model specification

In this thesis, we specify the noninvertible ARMA(P,Q) model as in Meitz and Saikkonen (2013),

$$y_t - a_{0,1}y_{t-1} - \cdots - a_{0,P}y_{t-P} = \varepsilon_t - b_{0,1}\varepsilon_{t+1} - \cdots - b_{0,Q}\varepsilon_{t+Q}, \tag{1.1}$$

where ε_t is non-Gaussian and iid. In this specification, models linear dependency on the current and future error terms is made precise. Model (1.1) can also be defined using the AR and MA polynomials as

$$a_0(B)y_t = b_0(B^{-1})\varepsilon_t,$$

where

$$a_0(B) = 1 - a_{0,1}B - \cdots - a_{0,P}B^P \quad \text{and}$$
$$b_0(B^{-1}) = 1 - b_{0,1}B^{-1} - \cdots - b_{0,P}B^{-Q},$$

where B is a backshift operator ($B^kX_t = X_{t-k}$, for all $k = \ldots, -1, 0, 1, \ldots$). The model has a stationary forward looking MA(∞) presentation, as long as polynomials $a_0(B)$ and $b_0(B^{-1})$ satisfy the root conditions

$$\begin{aligned} a_0(z) &\neq 0 \quad \text{for all} \quad |z| \leq 1 \quad \text{and} \\ b_0(z^{-1}) &\neq 0 \quad \text{for all} \quad |z^{-1}| \leq 1. \end{aligned} \tag{1.2}$$

Under these root conditions we have

$$y_t = \sum_{j=-Q}^{\infty} \psi_j \varepsilon_{t-j} \quad \text{and} \quad \varepsilon_t = \sum_{j=P}^{\infty} \pi_j y_{t+j}, \tag{1.3}$$

where ψ_j and π_j are geometrically decaying coefficients of the Laurent's series expansions of $a_0(B)^{-1}b_0(B^{-1})$ and $b_0(B^{-1})^{-1}a_0(B)$, respectively.

This specification differs slightly from those in Lii and Rosenblatt (1992) and Lii and Rosenblatt (1996). In their description of the model, the MA polynomial is $b_0(B)$, in contrast to the $b_0(B^{-1})$ in (1.1), and the root condition of the MA polynomial reads as $b_0(z) \neq 0$ for all $|z| \geq 1$. Both specifications span the same space of models, but there are certain benefits in our specification (see Chapter 2 and Meitz and Saikkonen, 2013, for more details).

1.2.2 Identification

As we already pointed out, non-Gaussianity is a necessary and sufficient condition for identifying the noninvertible model. In Rosenblatt (1985) this is illustrated by the following example. Let us consider an invertible ARMA(1,1) model

$$(1 - a_0 B)y_t = (1 - b_0 B)\varepsilon_t,$$

with Gaussian iid error term ε_t. Because y_t is obtained by a linear filter from a Gaussian sequence ε_t, it is Gaussian itself. Let $|\bar{z}_a| > 1$ and $|\bar{z}_b| > 1$ denote the roots of the AR and the MA polynomials, respectively. Process y_t has a spectral density

$$f_y(\omega) = \frac{\sigma_\varepsilon^2}{2\pi} \frac{|\bar{z}_b - e^{-i\omega}|^2}{|\bar{z}_a - e^{-i\omega}|^2} = \frac{\sigma_\varepsilon^2 |\bar{z}_b|^2}{2\pi} \frac{|\bar{z}_b^{-1} - e^{-i\omega}|^2}{|\bar{z}_a - e^{-i\omega}|^2}.$$

The r.h.s. is the spectral density of

$$(1 - a_0 B)y_t = (1 - b_0 B^{-1})\varepsilon_t^*,$$

a noninvertible ARMA model with $\text{Var}(\varepsilon_t^*) = |\bar{z}_b|^2\sigma_\varepsilon^2$. In the Gaussian case the whole probability structure of y_t is determined by the second moments of the process. The fact that these two processes have the same spectral densities implies that the models cannot be statistically identified. This is not, however, the case for other distributions, in general. The probability structures of the observations are different for different locations of the roots.

This example carries over ARMA(P,Q) models with any values of P and Q. The MA polynomial can always be factorized, and the roots can be switched from outside the unit circle to inside. In ARMA(1,1) case, there were two possible models with the same spectral densities, but for the MA polynomials of order Q, there are already 2^Q different models, one of them being strictly invertible, and one being strictly noninvertible. Rest of the models have roots both inside and outside of the unit circle.

1.2.3 Backward looking representations for the noninvertible ARMA model

Under the root conditions (1.2) it is easy to see that the process (1.1) is covariance-stationary. Wold's representation theorem says that all the covariance-stationary processes can be written as a sum of the history of the non-correlated shocks (and some nonrandom sequence). This is in contrast to the MA(∞) presentation (1.3), in which the process y_t is written in terms of the past, current, and future shocks. The question arises, what is the connection of the model (1.1) to the Wold's decomposition?

To see how these two are related, notice that the Wold's decomposition does not say that the presentation is linear in iid shocks, as in (1.3). We may take another point of view to the model (1.1) and consider it as a nonlinear model, that has a linear backward looking presentation in terms of uncorrelated error terms e_t,

$$y_t = \frac{b_0(B)}{a_0(B)} e_t, \quad \text{with} \quad e_t = \frac{b_0(B^{-1})}{b_0(B)} \varepsilon_t. \tag{1.4}$$

It is easy to verify that e_t is uncorrelated process by looking at its spectral density function

$$f_e(\omega) = \frac{\sigma_\varepsilon^2}{2\pi} \frac{b_0(e^{-2\pi i\omega}) b_0(e^{2\pi i\omega})}{b_0(e^{2\pi i}) b_0(e^{-2\pi i})} = \frac{\sigma_\varepsilon^2}{2\pi},$$

which is flat for all frequencies ω. Under the root conditions (1.2), the polynomials $a_0(B)$ and $b_0(B)$ clearly have their zeros outside the unit circle, so there exist coefficients $\{\psi^*_{0,j}\}_{j=0}^{\infty}$ of the power series expansion of $b_0(B)a_0(B)^{-1}$ such that $y_t = \sum_{j=0}^{\infty} \psi^*_{0,j} e_{t-j}$, a Wold decomposition of the noninvertible ARMA model (1.1).

The power of the Wold's representation theorem stems from the fact that ARMA processes are in general very good approximations to the infinite backward looking sums. In previous literature, the use of ARMA models has been justified by the fact, that no matter if the true data generating process is an ARMA process, by the Wold's representation theorem, it can be very closely approximated by one. But if the data was indeed generated by a noninvertible process, using this approximation would cause a loss of information in the modeling. Although e_t is an uncorrelated sequence, it is not independent, but an ARMA sequence itself. We would neglect the information implied by the ARMA structure of the error term, and we would effectively consider these errors as purely random error terms.

It is explained in Lanne, Meitz, and Saikkonen (2013) that e_t is a nonlinearly dependent sequence. For example, the squared values of e_t are linearly dependent. This is the reason why the noninvertible model is capable of controlling mild heteroskedasticity in the data.

This alternative representation (1.4) also illustrates a possible way of detecting a noninvertible model. By construction, the parameters in polynomials $a_0(B)$ and $b_0(B)$ can be estimated from the data by using any estimation method based on the second moments of the data. If the orders of the polynomials are correctly specified, the obtained residual series $\tilde{e}_t$ should be a white noise sequence. However, dependencies in the higher moments of this series would suggest that the underlying data generating process was noninvertible, and instead of (1.4) with white noise error e_t, we should perhaps consider estimating the noninvertible model (1.1).

In Chapter 4, we test a hypothesis, whether the process e_t in (1.4) is a good description for the quarterly U.S. stock return data, versus the alternative hypothesis of the data being generated by process y_t. Process e_t defines a so-called all-pass process (see Andrews, Davis, and Breidt, 2006; Breidt, Davis, and Trindade, 2001). This process generates data that are not correlated, but still dependent. This feature is desirable, since the lack of autocorrelation implies the independence of the observation for the standard ARMA models. Lanne et al. (2013) shows that this model is capable of controlling for mild heteroskedasticity in the data, and using the diagnostic tests developed in

Chapter 2, we show that it seems sufficient in capturing the correlation of the squared observations encountered in our stock return data. Another feature of this all-pass process is that it is predictable despite being an uncorrelated process. For this reason, the noninvertible model is prominent in answering the question of whether there is linear or nonlinear predictability in stock returns.

1.2.4 Predictability

Forecasting with the noninvertible ARMA model is a considerably more difficult task than with standard causal and invertible ARMA models. Meitz and Saikkonen (2013) illustrates this difficulty. Let us consider the conditional expectation of y_t given the observations up to the time $t-1$,

$$\begin{aligned} \mathrm{E}_{t-1}[y_t] =& a_{0,1}y_{t-1} + \cdots + a_{0,P}y_{t-P} \\ & + \mathrm{E}_{t-1}[\varepsilon_t] - b_{0,1}\mathrm{E}_{t-1}[\varepsilon_{t+1}] - \cdots - b_{0,Q}\mathrm{E}_{t-1}[\varepsilon_{t+Q}]. \end{aligned}$$

Although the error terms have mean zero, the conditional mean is not zero. In fact, the conditional mean of the errors is dependent on the history of y_t. This can be seen most clearly from (1.3), from where one can postulate that the error terms are correlated with both lagged and leading values of y_t. In Appendix A.1., in Lanne et al. (2013), there is a proof that conditional expectation $\mathrm{E}_{t-1}[\varepsilon_t]$ is not linear, and furthermore, it is not constant in general, but nonlinearly dependent on the history of y_t for any parameter values that satisfy the root conditions (1.2). We find this kind of nonlinear predictability in a wide variety of stock return portfolios in Chapter 4, which would go unnoticed if we based our conclusions on predictability only on autocorrelations.

As pointed out, although the best predictions (the conditional expectation $\mathrm{E}_{t-1}[y_t]$) are difficult to calculate in practice, the best linear predictions are much simpler. As pointed out in Meitz and Saikkonen (2013), because the error terms e_t are uncorrelated, linear predictions can be calculated using (1.4).

1.2.5 Illustrations

Figure 1.1 plots a simulated times series with $T = 250$ observations of the noninvertible ARMA process in (1.1) with $a_0 = 0.8$ and $b_0 = 0.2$. The error process is assumed to follow a re-scaled Student's t-distribution with degrees of freedom $\lambda_0 = 5$. The bottom panel in Figure 1.1 displays the sample autocorrelation coefficients of the series. On the left, we see that the series is auto-

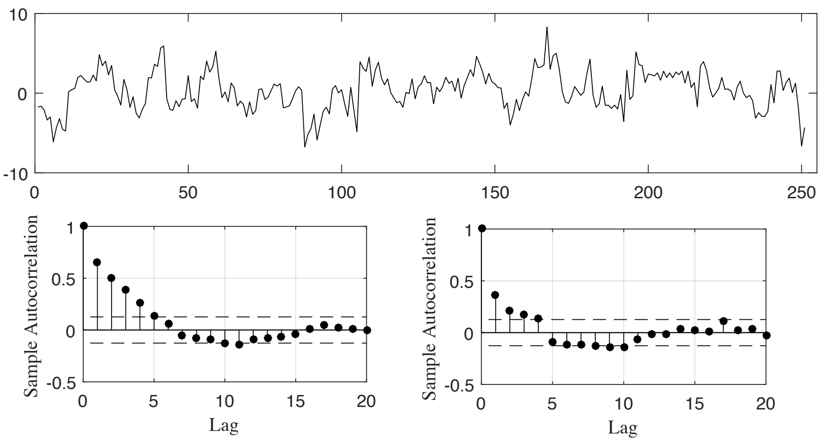

Figure 1.1: TOP: Simulation of $T = 250$ observations of the noninvertible ARMA(1,1) process (1.1) with $a_0 = 0.8$ and $b_0 = 0.2$. BOTTOM LEFT: Sample autocorrelation coefficients of the simulated series. BOTTOM RIGHT: Sample autocorrelation coefficients of the squared values of the simulated series.

correlated as we would expect, and the linear dependence decays smoothly to zero. On the right we have plotted the sample autocorrelation coefficients of the squared observations. This illustrates the nonlinear behavior of the noninvertible ARMA(1,1). The squared observations are autocorrelated and this implies the heteroskedasticity of the series. The pattern is similar to those of the ARCH processes, although it is rather mild.

Figure 1.2 depicts plotted values of an all-pass process e_t in (1.4) with $b_0 = 0.8$, and the distribution of the error term ε_t is the same as previously. The process is seemingly a weak white noise as the estimated sample autocorrelations are not significant at any lag. The squared observations, however, in the bottom right part of the figure, illustrates the nonlinear patterns of the all-pass process. Although the series is not correlated, the squared observations are, and the heteroskedasticity remains. This is obviously not the case with the conventional invertible ARMA models, in which the lack of autocorrelations automatically implies the independence of the observations. We encounter similar patterns in all of the subsequent chapters of this thesis. When we analyze financial time series data, we find that the return series are very mildly autocorrelated, but the heteroskedasticity is visible.

Baseline paths of the noninvertible ARMA processes are illustrated in Fig-

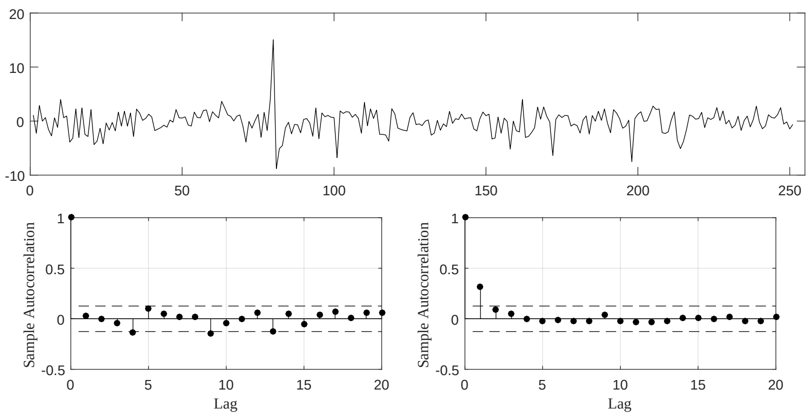

Figure 1.2: TOP: Simulation of $T = 250$ observations of the all-pass process e_t in (1.4) with $b_0 = 0.8$. BOTTOM LEFT: Sample autocorrelation coefficients of the simulated series. BOTTOM RIGHT: Sample autocorrelation coefficients of the squared values of the simulated series.

ure 1.3 together with baseline paths of the invertible ARMA(1,1) model with the same autocovariance structure. These plots highlight the economic importance of the identification of the noninvertible model. Recall the MA(∞) presentation (1.3) of the noninvertible ARMA process (1.1), $y_t = \sum_{j=-Q}^{\infty} \psi_j \varepsilon_{t-j}$. Using this, we have calculated the solid lines, which display the values of y_t for $t = -10, \ldots, 0, \ldots 10$, assuming that $\varepsilon_0 = -1$ and $\varepsilon_t = 0$ for all $t \neq 0$. The dashed lines, on the other hand, are the responses of the invertible ARMA process with the same autocovariance structure than the noninvertible model has. In the left panel we have set $a_0 = 0.5$ and $b_0 = 0.7$, and for the right panel we have $a_0 = 0.5$ and $b_0 = -0.7$.

If an econometrician fails to identify the noninvertibility of the process, and estimates an invertible ARMA model, he/she recovers the baseline path illustrated by the dashed line. As we can see, these paths may differ substantially, so the conclusions the econometrician draws from his/her analysis may be misleading. In the left panel of Figure (1.3), both of the paths have similar patterns, but the solid line reacts slightly stronger to the negative shock. On the other hand, it also recovers faster to zero. In the right panel, on the other hand, the initial reactions are of the opposite sign. If the data was generated by the noninvertible model, a negative shock at $t = 0$ would induce an initial positive reaction, after which the process turns negative and starts to recover towards its equilibrium. If the econometrician estimated an invertible model, he/she would observe the baseline path illustrated by the dashed line, and would conclude that a negative shock would induce an initial negative impact.

For example Leeper, Walker, and Yang (2013) show that misspecification of the noninvertibility may have serious consequences in policy analysis. They study an impact of a fiscal policy shock in an dynamic stochastic general equilibrium model assuming that the agents may foresee the forthcoming tax changes some periods in advance. After solving the model they find that, in their model economy, capital accumulation actually follows a noninvertible ARMA process. If a policymaker wishes to study the response of capital accumulation to a tax shock by using an ARMA model without identifying the noninvertibility of the true data generating process, he/she draws misleading conclusions and about the responses on his/her policy suggestions.

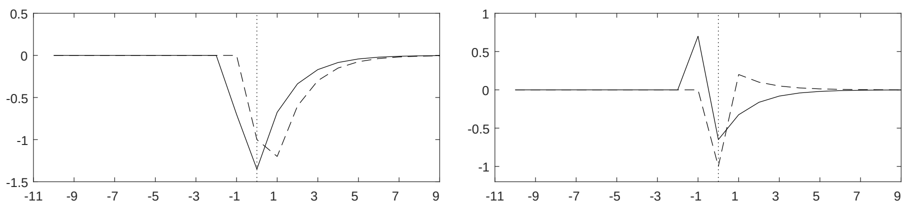

Figure 1.3: LEFT: Baseline paths for two different ARMA models. The solid line is for a noninvertible ARMA(1,1) $y_t = 0.5y_{t-1} + \varepsilon_t + 0.7\varepsilon_{t+1}$, and the dashed line is for an invertible ARMA(1,1) $y_t = 0.5y_{t-1} + \varepsilon_t + 0.7\varepsilon_{t-1}$, assuming that $\varepsilon_0 = -1$ and $\varepsilon_t = 0$ for all $t \neq 0$. RIGHT: Baseline paths for a noninvertible ARMA(1,1) $y_t = 0.5y_{t-1}\varepsilon_t - 0.7\varepsilon_{t+1}$ (solid) and invertible ARMA(1,1) $y_t = 0.5y_{t-1} - 0.7\varepsilon_{t-1}$ (dashed).

1.3 Noninvertibility and economic models

Economic models may result in noninvertible log-linearized solutions. These types of models can be found in the fields of asset pricing (Kasa, Walker, and Whiteman, 2014) and fiscal foresight (Leeper et al., 2013), and in news shocks models (Blanchard and Perotti, 2002; Forni and Gambetti, 2014), and permanent income models (Fernández-Villaverde, Rubio-Ramírez, Sargent, and Watson, 2007). What these economic models have in common is that the agents are allowed to foresee hints of the forthcoming shocks before they hit the economy. If this is the case, an econometrician who is trying to model the economy uses a smaller information set when performing the estimation than the model agents use. Information deficiency of this sort leads to the noninvertible solution (Hansen and Sargent, 1991). A comprehensive review of noninvertibility and economic models can be found in Alessi, Barigozzi, and Capasso (2011).

1.4 Noninvertibility in time series econometrics

In this thesis we concentrate on a specification of the noninvertible ARMA model by Meitz and Saikkonen (2013). This specification differs somewhat from the previous versions of noninvertible models. In this specification, the observations dependency of the future error terms is written down explicitly. Observations are linear combinations of a finite amount of past observations, and a finite amount of future error terms, in addition to the current error. Meitz and Saikkonen (2013) shows how to approximate the log-likelihood of the model, and they derive the asymptotic properties of the maximum likelihood estimator of the parameters.

From a time series perspective, noninvertible models, with non-causal models, possess some desirable features. Non-causal models may be seen as a good candidate for forecasting data, that can be considered forward-looking. This phenomena has been studied in Lanne, Nyberg, and Saarinen (2012) and Lanne, Luoto, and Saikkonen (2012), and in a multivariate setting in Nyberg and Saikkonen (2014). Gouriéroux and Zakoïan (2013) provides insight on how a simple non-causal model with a heavy-tailed error distribution can be used in modeling bubble phenomena in asset markets. Hecq, Lieb, and Telg (2016) illustrates how to identify non-causal models from their causal counterparts in small samples. They also study the time series of daily realized volatility of a set of financial assets, and conclude the presence of the non-causal component in that data. Hecq, Telg, and Lieb (2017) shows how seasonal adjustment of economic data induces non-causal dynamics in the adjusted series. Noninvertible models have been shown to provide a good fit for the financial data in Breidt et al. (2001) and in Lanne et al. (2013). Wu and Davis (2010) and Wu (2013) illustrate how to model financial data with noninvertible models accompanied by heavy-tailed error distribution.

1.5 Summary of the essays

1.5.1 Chapter 2: Residual based diagnostic tests for noninvertible ARMA models

In this chapter we propose two residual-based diagnostic tests for noninvertible ARMA models. The tests are analogous to the portmanteau tests developed by Box and Pierce (1970) and McLeod and Li (1983) in the conventional invertible case. We derive the asymptotic χ^2 distribution for the tests under the null of correctly specified model, and study the size and power properties in a Monte Carlo simulation study. An empirical application employing financial time series data points out the usefulness of noninvertible ARMA model in analyzing stock returns and the use of the proposed test statistics.

1.5.2 Chapter 3: Maximum likelihood estimation of a noninvertible ARMA model with α-stable errors

We study the properties of the maximum likelihood estimator of a noninvertible ARMA model with errors that follow an α-stable distribution and have infinite variance. To ensure the identification of the noninvertible ARMA model considered, we restrict the analysis to non-Gaussian distributions. Estimators of the autoregressive and moving average parameters are shown to be $n^{1/\alpha}$-consistent and to converge to a non-standard limiting distribution that is obtained as a maximizer of a certain random function. Estimators of the parameters in the α-stable distribution have the conventional $n^{1/2}$ rate of convergence. The finite sample properties of the estimators are studied in a simulation experiment, and an application to financial time series data from the New York Stock Exchange illustrates the applicability of the model.

1.5.3 Chapter 3: Nonlinear predictability of asset returns

For many theoretical asset pricing models, predictability follows as an implication of the risk aversion of agents. A closed form solutions for the next periods asset return depends on how the agents form their expectations about the future state of the world. By no means should this predictability be linear. First, we provide evidence of predictability of returns of U.S. stock portfolios and individual financial sector stocks using noninvertible ARMA model a and two-stage predictability testing procedure by Lanne et al. (2013). Second, we provide a straightforward extension to this procedure and allow for a larger model than noninvertible ARMA(1,1).

References

Alessi, L., M. Barigozzi, and M. Capasso (2011). Non-fundamentalness in structural econometric models: a review. *International Statistical Review 79*, 16–47.

Andrews, B., R. A. Davis, and F. J. Breidt (2006). Maximum likelihood estimation for all-pass time series models. *Journal of Multivariate Analysis 97*, 1638–1659.

Blanchard, O. and R. Perotti (2002). An empirical characterization of the dynamic effects of changes in government spending and taxes on output. *Quarterly Journal of Economics 117*, 1329–1368.

Box, G. and G. Jenkins (1970). *Time series analysis: forecasting and control*. Holden-Day series in time series analysis. Holden-Day.

Box, G. E. and D. A. Pierce (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association 65*, 1509–1526.

Breidt, F. J., R. A. Davis, and A. A. Trindade (2001). Least absolute deviation estimation for all-pass time series models. *Annals of Statistics 29*, 919–946.

Fernández-Villaverde, J., J. F. Rubio-Ramírez, T. J. Sargent, and M. W. Watson (2007). ABCs (and Ds) of understanding VARs. *American Economic Review 97*, 1021–1026.

Forni, M. and L. Gambetti (2014). Sufficient information in structural VARs. *Journal of Monetary Economics 66*, 124–136.

Gouriéroux, C. and J.-M. Zakoïan (2013). *Explosive bubble modelling by noncausal process*. CREST Discussion Papers.

Hansen, L. P. and T. J. Sargent (1991). Two difficulties in interpreting vector autoregressions. In L. P. Hansen, T. J. Sargent, J. Heaton, A. Marcet, and W. Roberds (Eds.), *Rational expectations econometrics*. Oxford: Westview Press Boulder, CO.

Hecq, A., L. Lieb, and S. Telg (2016). Identification of mixed causal-noncausal models in finite samples. *Annals of Economics and Statistics 123/124*, 307–331.

Hecq, A., S. Telg, and L. Lieb (2017). Do seasonal adjustments induce noncausal dynamics in inflation rates? *Econometrics 5*.

Kasa, K., T. B. Walker, and C. H. Whiteman (2014). Heterogeneous Beliefs and Tests of Present Value Models. *Review of Economic Studies 81*, 1137–1163.

Lanne, M., J. Luoto, and P. Saikkonen (2012). Optimal forecasting of noncausal autoregressive time series. *International Journal of Forecasting 28*, 623–631.

Lanne, M., M. Meitz, and P. Saikkonen (2013). Testing for linear and nonlinear predictability of stock returns. *Journal of Financial Econometrics 11*, 682–705.

Lanne, M., H. Nyberg, and E. Saarinen (2012). Does noncausality help in forecasting economic time series? *Economics Bulletin 32*, 2849–2859.

Leeper, E. M., T. B. Walker, and S.-C. S. Yang (2013). Fiscal foresight and information flows. *Econometrica 81*, 1115–1145.

Lii, K.-S. and M. Rosenblatt (1992). An approximate maximum likelihood estimation for non-gaussian non-minimum phase moving average processes. *Journal of Multivariate Analysis 43*, 272–299.

Lii, K.-S. and M. Rosenblatt (1996). Maximum likelihood estimation for non-gaussian nonminimum phase ARMA sequences. *Statistica Sinica 6*, 1–22.

Lof, M. (2013). Noncausality and asset pricing. *Studies in Nonlinear Dynamics and Econometrics 17*, 211–220.

McLeod, A. I. and W. K. Li (1983). Diagnostic checking arma time series models using squared-residual autocorrelations. *Journal of Time Series Analysis 4*, 269–273.

Meitz, M. and P. Saikkonen (2013). Maximum likelihood estimation of a non-invertible arma model with autoregressive conditional heteroskedasticity. *Journal of Multivariate Analysis 114*, 227–255.

Nyberg, H. and P. Saikkonen (2014). Forecasting with a noncausal VAR model. *Computational Statistics & Data Analysis 76*, 536–555.

Rosenblatt, M. (1985). *Stationary Sequences and Random Fields* (1 ed.). Birkhäuser Basel.

Singleton, K. J. (2009). *Empirical dynamic asset pricing: model specification and econometric assessment*. Princeton University Press.

Wu, R. (2013). M-estimation for general arma processes with infinite variance. *Scandinavian Journal of Statistics 40*, 571–591.

Wu, R. and R. A. Davis (2010). Least absolute deviation estimation for general autoregressive moving average time-series models. *Journal of Time Series Analysis 31*, 98–112.

2 Residual-based diagnostic tests for noninvertible ARMA models[1]

2.1 Introduction

Portmanteau tests as diagnostic tools in ARMA(P, Q) modeling of time series were made popular in a series of papers in the 1970s. The key idea of Box and Pierce (1970) was to derive the asymptotic distribution for a vector of m first empirical autocorrelation coefficients of an estimated residual of an ARMA model. Under the null hypothesis of a Gaussian, independent and identically distributed (iid) error term process of the ARMA model, and correctly specified estimated ARMA(P, Q) model, they showed that this vector is asymptotically normally distributed. Using this finding, they formulated the Pox-Pierce Q statistic, which was asymptotically χ^2_{m-P-Q} distributed under the null hypothesis. A key insight in deriving the limiting distribution was how to incorporate the estimation uncertainty in the residuals into the asymptotic result. Ljung and Box (1978) revisited these results, and proposed

[1]I am grateful for the guidance and suggestions of Markku Lanne, Mika Meitz and Pentti Saikkonen. I am also thankful to Tim Bollerslev and all the participants at Financial Time Series Econometrics lunch group at Duke University; Henri Nyberg, and the participants at HECER Econometrics Workshops; the participants at the 9^{th} Nordic Econometric Meeting; and Christian Gouriéroux and Jean-Michel Zakoïan and the participants at SoFiE Financial Econometrics summer school 2016, for useful comments. Detailed comments of Beth Andrews and Alain Hecq are very much appreciated. Financial support from Academy of Finland, Yrjö Jahnsson Foundation and OP-Pohjola Group Research Foundation is gratefully acknowledged. An earlier version of this chapter has been published as *HECER Discussion Paper No. 416*.

the Ljung-Box Q statistic to improve the small sample properties of the test.

A similar line of reasoning was used in McLeod and Li (1983) to derive the McLeod-Li Q test for checking autocorrelation in the squared residuals of an ARMA model. They noticed that the estimation uncertainty was not an issue in this case. The asymptotic distribution of the empirical autocorrelation coefficients of the estimated squared residuals was not affected by the distribution of the estimated parameters of the ARMA(P, Q) model, and the limiting distribution of their test was χ^2_m. Since these seminal findings, it has become standard practice to reject the estimated ARMA(P, Q) model at the confidence level α if the aforementioned test statistics exceed the $(1-\alpha)$-quantile of the corresponding χ^2-distributions.

Development of the estimation theory for new kinds of linear ARMA models and nonlinear time series models have inspired theory for new tests in the fashion of the previous papers. Romano and Thombs (1996) showed that it is important to take the underlying assumptions of the ARMA models into account. The critical values of the standard tests might be misleading if the iid assumption fails and the error term is merely a martingale difference or weak white noise, for example. Li and McLeod (1988) showed how to modify the test in the case of maximum likelihood estimation with non-Gaussian errors. Francq, Roy, and Zakoïan (2012) focused on more robust tests, which assume only weak white noise error terms, allowing for error processes that are not martingale differences. Lin and McLeod (2008) considered models with stable distributed innovations allowing for infinite variance and possibly infinite mean. Many other formulations have been suggested to improve the tests' small sample properties and adapt them for different assumptions of the ARMA models, see for example, Monti (1994); Peña and Rodríguez (2002); Chen and Deo (2004); Lobato (2001); Lobato, Nankervis, and Savin (2001, 2002).

In this paper, we concentrate on the consequences of the noninvertibility of the ARMA model for the portmanteau tests. Because non-Gaussianity is essential for the identification of the noninvertible ARMA model, our paper is closely related to Li and McLeod (1988). Our paper overlaps with, but is not nested by Francq et al. (2012). In principle, their test can be used for model selection among noninvertible models. However, it does not allow for the efficient ML estimation of the parameters. We, on the other hand, build on the ML estimation theory for the noninvertible processes, introduced by Meitz and Saikkonen (2013). Thus, our methods allows for more efficient parameter estimation of assumable noninvertible and non-Gaussian ARMA processes and draws the critical values using the actual dependence structure of the data. Our paper is also related to the work of Cui, Fisher, and Wu (2014) who study autocorrelation tests for noncausal autoregressive processes with stable distributed errors.

The usefulness of noninvertible time series models in economic research was first pointed out by Hansen and Sargent (1981, 1991). They show that the information deficiency of the econometrician leads to a noninvertible (or nonfundamental) solution of the model. More recently, nonfundamentalness has arisen in asset pricing models (Kasa, Walker, and Whiteman, 2014), fiscal foresight models (Leeper, Walker, and Yang, 2013), news shocks models (Blanchard and Perotti, 2002; Forni and Gambetti, 2014), and permanent income models (Fernández-Villaverde, Rubio-Ramírez, Sargent, and Watson, 2007). It is common to all of these models that they have a noninvertible linearized solution. A comprehensive survey on noninvertibility in economic theory can be found in Alessi, Barigozzi, and Capasso (2011).

Empirical evidence of the good fit of the noninvertible ARMA model to economic time series data have been provided by Andrews, Calder, and Davis (2009), Breidt, Davis, and Trindade (2001), and Huang and Pawitan (2000). Lanne and Saikkonen (2013) point out that noninvertible models are potentially capable of capturing the nonlinearities in the time series as they are driven by the iid error terms in a nonlinear fashion. For example, these models are shown to control for mild heteroskedasticity, commonly encountered in financial time series. They are also capable of producing time series that are at most very mildly autocorrelated but still nonlinearly dependent. These nonlinearities cannot be controlled by a Gaussian causal and invertible ARMA models, as the lack of autocorrelation automatically implies the independence of the observations.

Noninvertible ARMA(P, Q) model have a linear representation in reversed time. That is, the model is linear in terms of the P past observations and the current and Q future errors. This forward-looking feature has been shown useful in modeling and forecasting locally explosive bubbles in commodity markets, as shown in Gouriéroux and Zakoïan (2017), Cavaliere, Nielsen, and Rahbek (2018), and Friés and Zakoïan (2019), using mixed causal-noncausal models, which share a somewhat similar forward-looking structure than our noninvertible model.

We build our empirical application on the work by Lanne, Meitz, and Saikkonen (2013), who propose a two-step procedure for testing the predictability in financial time series. This test is based on the properties of the noninvertible ARMA model, and it is crucial to test if the model fits the data well. We conclude that noninvertible ARMA$(1, 1)$ provides a good fit for stock return data, so the findings of nonlinear predictability are credible from this perspective.

The rest of the paper is organized as follows. Section 2.2 describes the noninvertible ARMA model in detail and briefly discusses its maximum likelihood estimation. Section 2.3 introduces the test statistics considered and derives their asymptotic properties. In Section 2.4 we conduct a Monte Carlo experiment to study the small sample properties of the tests. Section 2.5 provides an empirical example using financial time series data used by Lanne et al. (2013). Section 2.6 concludes. High-level assumptions are left for the appendices to ease reading, as well as some intermediate results and lemmas used in the proofs of the main results. The supplementary appendix contains detailed proofs for the intermediate results.

A few notational conventions are given. Almost sure (a.s.) convergence, convergence in probability, and in distribution are denoted by $\overset{a.s.}{\to}$, $\overset{p}{\to}$, and $\overset{d}{\to}$, respectively. We use $\overset{as}{\sim}$ when the left and the right hand sides have the same asymptotic limiting distribution. All vectors are column vectors unless otherwise indicated. That is, $(x_1, \ldots, x_h)$ is a column vector where the h elements are either scalars or column vectors. The L_r-norm is denoted by $\| \cdot \|_r$, meaning that for any random variable x, $\|x\| = (\mathrm{E}[|x|^r])^{1/r}$, where $r > 0$. Identity matrix of size m is denoted by $\boldsymbol{I}_m$ and $(m \times n)$ matrix of zeros is denoted by $\mathbf{0}_{m \times n}$.

2.2 The noninvertible ARMA model

The maximum likelihood estimation of noninvertible (AR)MA model has been studied, among others, by Lii and Rosenblatt (1992, 1996).[2] We study the purely causal and noninvertible ARMA(P, Q) process by Meitz and Saikkonen (2013):

$$a_0(B) y_t = b_0(B^{-1}) \varepsilon_t, \tag{2.1}$$

where B denotes the backward shift operator, $B^k x_t = x_{t-k}$, for $k \in \mathbb{Z}$, $a_0(z) = 1 - a_{0,1} z - \cdots - a_{0,P} z^P$ is an autoregressive (AR) polynomial of order P and

[2] Breidt et al. (2001) and Andrews, Davis, and Breidt (2006) provide estimation theory for an important special case of the noninvertible ARMA model, the so called all-pass model.

has its roots outside the unit circle:

$$a_0(z) \neq 0 \quad \text{for all } |z| \leq 1.$$

Moving average (MA) polynomial $b_0(z^{-1}) = 1 - b_{0,1}z^{-1} - \cdots - b_{0,Q}z^{-Q}$ is of order Q, and its roots lie outside the unit circle:

$$b_0(z^{-1}) \neq 0 \quad \text{for all } |z^{-1}| \leq 1.$$

Error term $\varepsilon_t = \sigma_0\eta_t$ is assumed non-Gaussian and iid, with zero mean and finite variance: $\mathrm{E}[\varepsilon_t] = 0$ and $\mathrm{E}[\varepsilon_t^2] = \sigma_0^2 < \infty$. It is assumed that η_t has a symmetric distribution with density function $f_\eta(x; \boldsymbol{\lambda}_0)$ with a $(d \times 1)$ parameter vector $\boldsymbol{\lambda}_0$.

This presentation differs somewhat from, for example, that in Lii and Rosenblatt (1996), where the MA polynomial is written as $b_0(B)$ (in contrast to $b_0(B^{-1})$). In their presentation, the MA polynomial is written in terms of the current and the past error terms, and the roots of the MA polynomial are allowed to situate inside or outside the unit circle. The pure noninvertibility would imply that all the roots were inside the unit circle, and if this were the case, the set of models would coincide in the current paper and in Lii and Rosenblatt (1996).[3] The difference is that in our presentation, the dependence of y_t on the future error terms ε_t is written explicitly. As pointed out by Meitz and Saikkonen (2013), although all the purely noninvertible ARMA(P, Q) models can be presented in either way, there are some consequences in the choice of presentation. The derivation of the approximate likelihood requires some care, as the observations are dependent on the future errors (see the derivations in Section 3.1. in Meitz and Saikkonen, 2013). A benefit of our presentation is that the likelihood function does not contain a term like $\log|b_Q|$ (Lii and Rosenblatt, 1996, p. 8), which allows for a straightforward test for an unknown order of the MA polynomial.

Non-Gaussianity of the error term is assumed for identification of the model. For each causal and invertible Gaussian ARMA model, there is always a noninvertible model with exactly the same second order properties so that the invertible and noninvertible ARMA models cannot be distinguished from each other (for a thorough discussion, see Rosenblatt (2012), Chapter 2).

[3]Consider an MA(1) model $y_t = (1 - aB)\zeta_t$ with $|a| > 1$ and ζ_t an iid sequence. Another way of writing this model is $y_t = (1 - a^{-1}B^{-1})\varepsilon_t$, where $\varepsilon_t = -aB\zeta_t$ is another iid sequence. The former is the presentation of Lii and Rosenblatt (1996) and the latter is in the form of Meitz and Saikkonen (2013). This idea generalizes easily for all the MA(Q) polynomials.

Model (2.1) has an MA(∞) representation in terms of Q future, the present, and infinite history of the error terms ε_t. It also has an AR(∞) representation in terms of P lagged, the present, and infinite future of the observations y_t,

$$y_t = \sum_{j=-Q}^{\infty} \psi_{0,j}\varepsilon_{t-j} \quad \text{and} \quad \varepsilon_t = \sum_{j=-P}^{\infty} \phi_{0,j} y_{t+j}.$$

The coefficients $\psi_{0,j}$ and $\phi_{0,j}$ are geometrically decaying coefficients of the Laurent series expansions of $a_0(z)^{-1}b_0(z^{-1})$ and $a_0(z)b_0(z^{-1})^{-1}$, respectively.[4]

For all the parameter values $\boldsymbol{\theta} = (a_1, \ldots, a_P, b_1, \ldots, b_Q, \sigma, \boldsymbol{\lambda}) \in \Theta$, where Θ is a permissible parameter space defined in Assumption 2 in Appendix A, we define

$$u_t(\boldsymbol{\theta}) = \frac{a(B)}{b(B^{-1})} y_t = \sum_{j=-P}^{\infty} \phi_j y_{t+j}, \tag{2.2}$$

where $a(z) = 1 - a_1 z - \cdots - a_P z^P$ and $b(z) = 1 - b_1 z - \cdots - b_Q z^Q$. This sum is well defined for all $\boldsymbol{\theta} \in \Theta$ and the coefficients ϕ_j decay geometrically as $j \to \infty$. Because the infinite future of y_t is not observable, this is an unfeasible way to approximate the error terms.

Assuming that $\{y_t\}_{t=1-P}^{T}$ are observed, the feasible counterpart of $u_t(\boldsymbol{\theta})$, say $\tilde{u}_t(\boldsymbol{\theta})$, is obtained using these observations and some initial values $\tilde{u}_{T+1}(\boldsymbol{\theta}) = \tilde{u}_{T+Q}(\boldsymbol{\theta}) = 0$. For $t = T, \ldots, 1$,

$$\tilde{u}_t(\boldsymbol{\theta}) = y_t - a_1 y_{t-1} - \cdots - a_P y_{t-P} + b_1 \tilde{u}_{t+1}(\boldsymbol{\theta}) + \cdots + b_{t+Q} \tilde{u}_{t+Q}(\boldsymbol{\theta}). \tag{2.3}$$

Regarding parameter estimation, Meitz and Saikkonen (2013) discuss the maximum likelihood estimation of Model (2.1) with an error term assumed to follow an ARCH process. Our model is thus a simplified version of theirs and the asymptotic properties of the ML estimator are obtained in a very similar fashion as in their paper. The properties are listed in Proposition 1 below. Let $L_T(\boldsymbol{\theta})$ denote an approximation of the log-likelihood function of the model (see Meitz and Saikkonen, 2013, for details),

$$L_T(\boldsymbol{\theta}) = T^{-1} \sum_{t=1}^{T} l_t(\boldsymbol{\theta}) \quad \text{with} \quad l_t(\boldsymbol{\theta}) = \log f_\eta(\sigma^{-1} u_t(\boldsymbol{\theta}); \boldsymbol{\lambda}) - \log \sigma, \tag{2.4}$$

[4] See Lemmas A.1. and A.2. in Meitz and Saikkonen (2013) for a thorough discussion of these series presentations.

and let $L_{\boldsymbol{\theta},T}(\boldsymbol{\theta})$ and $L_{\boldsymbol{\theta\theta},T}(\boldsymbol{\theta})$ denote the first and second order derivatives of the log-likelihood with respect to the parameter vector $\boldsymbol{\theta}$. Estimation and statistical inference is based on feasible versions of these quantities, denoted by $\tilde{L}_T(\boldsymbol{\theta})$, $\tilde{L}_{\boldsymbol{\theta},T}(\boldsymbol{\theta})$ and $\tilde{L}_{\boldsymbol{\theta\theta},T}(\boldsymbol{\theta})$, which are obtained by replacing $u_t(\boldsymbol{\theta})$ by $\tilde{u}_t(\boldsymbol{\theta})$ in the log-likelihood functions (the exact expressions can be found in Appendix 2.C).

The following proposition contains the conventional properties of a (local) maximum likelihood estimator. We omit the proof for brevity, but the arguments are very similar to those in Meitz and Saikkonen (2013).

Proposition 1. *Under Assumptions 1 (a), 2 and 3 in Appendix A,*

1. $\lim_{T\to\infty} Cov(T^{1/2}L_{\boldsymbol{\theta},T}(\boldsymbol{\theta}_0)) = \boldsymbol{\mathcal{I}}$, *where* $\boldsymbol{\mathcal{I}}$ *is positive definite and* $\boldsymbol{\mathcal{I}} = -E[l_{\boldsymbol{\theta\theta},t}(\boldsymbol{\theta}_0)]$,
2. $T^{1/2}L_{\boldsymbol{\theta},T}(\boldsymbol{\theta}_0) \xrightarrow{d} N(0, \boldsymbol{\mathcal{I}})$,
3. $\sup_{\boldsymbol{\theta}\in\Theta_0} |L_{\boldsymbol{\theta\theta},T}(\boldsymbol{\theta}) - \boldsymbol{\mathcal{J}}(\boldsymbol{\theta})| \to \mathbf{0}$ *a.s. as* $T \to \infty$, *where* $\boldsymbol{\mathcal{J}}(\boldsymbol{\theta}) = E[l_{\boldsymbol{\theta\theta},t}(\boldsymbol{\theta})]$ *is finite and continuous at* $\boldsymbol{\theta}_0$, *and* Θ_0 *is defined in Assumption 2 in Appendix A.*
4. *there exists a sequence of solutions* $\tilde{\boldsymbol{\theta}}_T$ *to likelihood equations* $\tilde{L}_{\boldsymbol{\theta},T}(\boldsymbol{\theta}) = 0$ *s.t.* $T^{1/2}(\tilde{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \xrightarrow{d} N(0, \boldsymbol{\mathcal{I}}^{-1})$, *and*
5. *there is a consistent estimator for the asymptotic covariance matrix given by the inverse of the Hessian,* $-\tilde{L}^{-1}_{\boldsymbol{\theta\theta},T}(\tilde{\boldsymbol{\theta}}_T) \to \boldsymbol{\mathcal{I}}^{-1}$ *a.s. as* $T \to \infty$.

These asymptotic properties of the ML estimators are the main ingredients for the asymptotic behavior of the test statistics we derive in the next section.

2.3 Diagnostic tests

We now introduce the main test statistics considered in this paper. The modified test statistics for residual autocorrelation and autocorrelation in squared residuals are[5]

$$Q_{ac} = T\tilde{\boldsymbol{\rho}}'_{ac,m}\tilde{\boldsymbol{\Omega}}^{-1}_{ac}\tilde{\boldsymbol{\rho}}_{ac,m} \quad \text{and} \quad Q_{hs} = T\tilde{\boldsymbol{\rho}}'_{hs,m}\tilde{\boldsymbol{\rho}}_{hs,m},$$

[5] All the quantities Q_{ac}, Q_{hs}, $\tilde{\boldsymbol{\rho}}_{ac,m}$, $\tilde{\boldsymbol{\rho}}_{hs,m}$ and so on are dependent on the sample size T, but for the sake of brevity, we suppress this dependence in our notation.

with $\tilde{\boldsymbol{\rho}}_{ac,m} = \tilde{\boldsymbol{\rho}}_{ac,m}(\tilde{\boldsymbol{\theta}}_T) = (\tilde{\rho}_{1,ac}(\tilde{\boldsymbol{\theta}}_T), \ldots, \tilde{\rho}_{m,ac}(\tilde{\boldsymbol{\theta}}_T))$, an m-vector of empirical autocorrelations of $\tilde{u}_t(\tilde{\boldsymbol{\theta}}_T)$. Respectively, $\tilde{\boldsymbol{\rho}}_{hs,m} = \tilde{\boldsymbol{\rho}}_{hs,m}(\tilde{\boldsymbol{\theta}}_T) = (\tilde{\rho}_{1,hs}(\tilde{\boldsymbol{\theta}}_T), \ldots$ $\ldots, \tilde{\rho}_{m,hs}(\tilde{\boldsymbol{\theta}}_T))$ is an m-vector of empirical autocorrelations of $\tilde{u}_t(\tilde{\boldsymbol{\theta}}_T)^2$. The i^{th} autocorrelation coefficient is calculated as $\tilde{\rho}_{i,ac}(\boldsymbol{\theta}) = \tilde{\gamma}_{i,ac}(\boldsymbol{\theta})/\tilde{\gamma}_{0,ac}(\boldsymbol{\theta})$ and $\tilde{\rho}_{i,hs}(\boldsymbol{\theta}) = \tilde{\gamma}_{i,hs}(\boldsymbol{\theta})/\tilde{\gamma}_{0,hs}(\boldsymbol{\theta})$ with

$$\tilde{\gamma}_{i,ac}(\boldsymbol{\theta}) = (T-i)^{-1} \sum_{t=i+1}^{T} \tilde{u}_t(\boldsymbol{\theta}_T)\tilde{u}_{t-i}(\boldsymbol{\theta}) \quad \text{and}$$
$$\tilde{\gamma}_{i,hs}(\boldsymbol{\theta}) = (T-i)^{-1} \sum_{t=i+1}^{T} (\tilde{u}_t(\boldsymbol{\theta}_T)^2 - \sigma^2)(\tilde{u}_{t-i}(\boldsymbol{\theta})^2 - \sigma^2).$$

The m-vectors of autocovariances of $\tilde{u}_t(\tilde{\boldsymbol{\theta}}_T)$ and $\tilde{u}_t(\tilde{\boldsymbol{\theta}}_T)^2$ are denoted by $\tilde{\boldsymbol{\gamma}}_{ac,m} = \tilde{\boldsymbol{\gamma}}_{ac,m}(\tilde{\boldsymbol{\theta}}_T) = (\tilde{\gamma}_{1,ac}(\tilde{\boldsymbol{\theta}}_T), \ldots, \tilde{\gamma}_{m,ac}(\tilde{\boldsymbol{\theta}}_T))$ and $\tilde{\boldsymbol{\gamma}}_{hs,m} = \tilde{\boldsymbol{\gamma}}_{hs,m}(\tilde{\boldsymbol{\theta}}_T) = (\tilde{\gamma}_{1,hs}(\tilde{\boldsymbol{\theta}}_T), \ldots$ $\ldots, \tilde{\gamma}_{m,hs}(\tilde{\boldsymbol{\theta}}_T))$, respectively. Vectors $\tilde{\boldsymbol{\rho}}_{ac,m}$ and $\tilde{\boldsymbol{\rho}}_{hs,m}$ are feasible statistics, since they are calculated using the observable quantities $\tilde{u}_t(\tilde{\boldsymbol{\theta}}_T)$ and $\tilde{\sigma}_T^2$. The dependence from the parameter value is suppressed in the notation to save space whenever we consider these feasible quantities at $\tilde{\boldsymbol{\theta}}_T$. Positive definite $(m \times m)$ matrix $\tilde{\boldsymbol{\Omega}}_{ac}^{-1}$ is a consistent estimator for the inverse of the limiting covariance matrix of $\tilde{\boldsymbol{\rho}}_{ac,m}$. The exact form will be given in the next subsection.

For the theoretical considerations that follow, it will be convenient to also consider unfeasible quantities

$$\boldsymbol{\rho}_{ac,m}(\boldsymbol{\theta}) = (\rho_{ac,1}(\boldsymbol{\theta}), \ldots, \rho_{ac,m}(\boldsymbol{\theta})), \quad \boldsymbol{\rho}_{hs,m}(\boldsymbol{\theta}) = (\rho_{hs,1}(\boldsymbol{\theta}), \ldots, \rho_{hs,m}(\boldsymbol{\theta})),$$
$$\boldsymbol{\gamma}_{ac,m}(\boldsymbol{\theta}) = (\gamma_{ac,1}(\boldsymbol{\theta}), \ldots, \gamma_{ac,m}(\boldsymbol{\theta})), \quad \text{and} \quad \boldsymbol{\gamma}_{hs,m} = (\gamma_{hs,1}(\boldsymbol{\theta}), \ldots, \gamma_{hs,m}(\boldsymbol{\theta})),$$

where $\rho_{i,ac}(\boldsymbol{\theta}) = \gamma_{i,ac}(\boldsymbol{\theta})/\gamma_{0,ac}(\boldsymbol{\theta})$ and $\rho_{i,hs}(\boldsymbol{\theta}) = \gamma_{i,hs}(\boldsymbol{\theta})/\gamma_{0,hs}(\boldsymbol{\theta})$. These quantities are unfeasible, since they are calculated using the unobservable variables $u_t(\boldsymbol{\theta})$,

$$\gamma_{i,ac}(\boldsymbol{\theta}) = (T-i)^{-1} \sum_{t=i+1}^{T} u_t(\boldsymbol{\theta})u_{t-i}(\boldsymbol{\theta}) \quad \text{and}$$
$$\gamma_{i,hs}(\boldsymbol{\theta}) = (T-i)^{-1} \sum_{t=i+1}^{T} (u_t(\boldsymbol{\theta}) - \sigma^2)(u_{t-i}(\boldsymbol{\theta})^2 - \sigma^2).$$

Note that $\gamma_{i,ac}(\boldsymbol{\theta}_0) = \gamma_{i,ac}$ and $\gamma_{i,hs}(\boldsymbol{\theta}_0) = \gamma_{i,hs}$ denotes the (unobservable) sample autocorrelations of ε_t and ε_t^2, as $u_t(\boldsymbol{\theta}_0) = \varepsilon_t$.

The form of the Q_{ac} test is the same as in Li and McLeod (1988). The covariance matrix between the vectors of empirical autocorrelation coefficients is needed to adjust for the non-Gaussianity of the data (we will return to this issue in Section 2.3.3). Remember, however, that the test statistic must be calculated using the feasible quantities that are recursively solved top-down. After one obtains the feasible residuals of the estimated noninvertible ARMA model in the appropriate manner, the Q_{hs} has exactly the same structure and asymptotics as the McLeod-Li test.

The null hypotheses under which the asymptotic properties of Q_{ac} and Q_{hs} are derived are slightly different for these tests, so we give two different null hypotheses (the Assumptions 1 - 3 are given in the Appendix),

$H_{0,ac}$: y_t admits model (2.1) and Assumptions 1, 2, and 3 (a).

$H_{0,hs}$: y_t satisfies $H_{0,ac}$ and Assumption 3 (b).

The difference between these null hypotheses is that the asymptotics of the Q_{ac} test are derived under the assumption of finite fourth moments, whereas the asymptotics of the Q_{hs} are shown under the assumption of finite eight moments. For example, if the Student's t-distribution is assumed for the error process, this assumption says that the degrees-of-freedom parameter is larger than eight. This assumption is admittedly strong, but a standard in the related literature. We will comment on it again in the simulation study in section 2.4.

Tests should have power against a variety of alternatives. For example, y_t might admit model (2.1) with orders $P^\star$ and $Q^\star$, where $P^\star > P$ or $Q^\star > Q$. This alternative is the main reason for the Q_{ac} test, as selecting too few lags or leads for the estimated model would result in autocorrelation in the estimated residuals. Of course there are many models that do not admit model (2.1), for example ARMA models with ARCH-type heteroskedasticity, which should be detected by these tests. Yet another example of the usefulness of Q_{hs} is an alternative, where the true model is a causal and invertible ARMA(P,Q). If the noninvertible model is fitted, the residuals should not exhibit autocorrelation, but Q_{hs} should be able to detect the autocorrelation in the squared residuals.[6]

[6]It was pointed out by pre-examiner, that heteroskedasticity in the residuals of the estimated model (2.1) may also be an implication of non-causality of the true DGP. We acknowledge this possibility, and state that our Q_{hs} test should be able to detect this sort of a misspecification as well, although this heteroskedasticity might be mild. We will not, however, elaborate this any further, since as such, the asymptotic properties of our tests are derived under strict causality of the AR polynomial, and even if the heteroskedasticity was detected, these tests could not be used to study the residuals of the estimated non-causal ARMA models. The tests for non-causal

Before we go into the asymptotics of our new tests, we wish to address the overlapping of the current paper and that by Francq et al. (2012). Our diagnostic tests efficiently use the entire probability structure of the noninvertible ARMA process y_t, as they rely on the asymptotic properties of the ML estimators of the model. It is well known, that the process (2.1) has an invertible and causal ARMA(P, Q) presentation in terms of a dependent weak white noise process[7]: $a_0(B)y_t = b_0(B)\zeta_t$, with $\zeta_t = b_0(B^{-1})b_0(B)^{-1}\varepsilon_t$. If one fits a causal and invertible ARMA(P, Q) model to the data y_t by using the second moments of the data, the resulting residuals should be approximately weak white noise. Francq et al. (2012) show how to test for the hypothesis that the resulting series is uncorrelated. If the null of no-autocorrelation cannot be rejected, then the estimated ARMA(P, Q) model seems an adequate description of the data. However, one must use some additional information in order to identify whether the model is purely invertible or noninvertible. This result also complements our results. If one cannot reject the weak white noise hypothesis, but there are visible nonlinear dependencies in the residuals, it encourages one to estimate the noninvertible ARMA(P, Q) model with the same P and Q as in the first step, and execute the new Q_{ac} and Q_{hs} tests. In this aspect, Francq et al. (2012) provides the means to select the orders P and Q without the need to estimatethe noninvertible ARMA model. However, once the orders have been selected, ML estimation of the noninvertible model improves the efficiency of the estimation, and our Q_{ac} and Q_{hs} test statistics can be used to investigatethe iid property of the residuals of that model more thoroughly.

Another branch of the literature concentrates on testing the iid hypothesis directly using the generalized spectral density approach by Hong (1999) (see also Hong and Lee, 2003). These tests are potentially useful in detecting the nonlinear dependencies in the residuals of the estimated invertible ARMA model, if the data is generated by the noninvertible model. However, it is not clear how these tests behave if one wishes to test for the independence of the residuals of the noninvertible ARMA model.[8]

ARMA processes are left for a potential line for future research. For noncausal AR processes with stable distributed errors, autocorrelation test is provided by Cui et al. (2014).

[7]It is clear that the process ζ_t is white noise as its spectral density is constant: $f_\zeta(\omega) = \frac{\sigma^2}{2\pi}\left|\frac{b_0(e^{-i\omega})}{b_0(e^{i\omega})}\right|^2 = \frac{\sigma^2}{2\pi}$.

[8]For example Velasco and Lobato (2018) propose, that the higher order spectral densities can be used to formulate a minimum distance estimator for estimating a possibly non-causal and/or noninvertible ARMA model.

2.3.1 Limiting distribution of Q_{ac}

We begin by establishing a limiting joint distribution result for an unobservable vector $T^{1/2}(\tilde{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0, \gamma_{ac,m})$. Although this result is not applicable as such in the diagnostic testing, it is the key to establishing the asymptotic properties for the feasible test statistic. The result is characterized by the model parameters and the asymptotic covariance matrix $\mathcal{I}$, so, in principle, the limiting distribution can be easily estimated by using consistent estimators for these parameters, given in Proposition 1.

To this end, let us denote $\boldsymbol{\Sigma}_{\gamma_{ac}} = \sigma_0^4 \boldsymbol{I}_{m\times m}$ and an $(m \times (P+Q))$ matrix

$$\boldsymbol{\Lambda}_m = \begin{pmatrix} \psi_{0,0}^{(a)} & \cdots & \psi_{0,1-P}^{(a)} & -\psi_{0,0}^{(b)} & \cdots & -\psi_{0,1-Q}^{(b)} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \psi_{0,m-1}^{(a)} & \cdots & \psi_{0,m-P}^{(a)} & -\psi_{0,m-1}^{(b)} & \cdots & -\psi_{0,m-Q}^{(b)} \end{pmatrix}, \tag{2.5}$$

where by convention, $\psi_{0,j}^{(a)} = \psi_{0,j}^{(b)} = 0$ when $j < 0$.

Lemma 1. *Under H_{ac}, for $P, Q > 0$, $T^{1/2}(\tilde{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0, \gamma_{ac,m}) \xrightarrow{d} N(0, \boldsymbol{\Sigma}_{\tilde{\boldsymbol{\theta}},\gamma_{ac}})$, where*

$$\boldsymbol{\Sigma}_{\tilde{\boldsymbol{\theta}},\gamma_{ac}} = \begin{pmatrix} \mathcal{I}^{-1} & \mathcal{I}^{-1}\boldsymbol{\Sigma}_{l,\gamma_{ac}}' \\ \boldsymbol{\Sigma}_{l,\gamma_{ac}}\mathcal{I}^{-1} & \boldsymbol{\Sigma}_{\gamma_{ac}} \end{pmatrix}, \quad \textit{with} \quad \boldsymbol{\Sigma}_{l,\gamma_{ac}} = \sigma_0^2 \begin{pmatrix} \boldsymbol{\Lambda}_m & \mathbf{0}_{m\times(1+d)} \end{pmatrix}.$$

Proof. Using mean value expansion of the (feasible) score function around the true parameter value $\boldsymbol{\theta}_0$ gives us

$$T^{1/2}\begin{pmatrix} \tilde{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0 \\ \gamma_{ac} \end{pmatrix} = T^{1/2}\begin{pmatrix} -L_{\boldsymbol{\theta\theta},T}^{-1}(\boldsymbol{\theta}_0)L_{\boldsymbol{\theta},T}(\boldsymbol{\theta}_0) \\ \gamma_{ac} \end{pmatrix} + T^{1/2}\begin{pmatrix} R_{1,T} \\ \mathbf{0}_{m\times 1} \end{pmatrix}, \tag{2.6}$$

where

$$\begin{aligned} R_{1,T} = &-(\tilde{L}_{\boldsymbol{\theta\theta},T}^{-1}(\bar{\boldsymbol{\theta}}_T) - L_{\boldsymbol{\theta\theta},T}^{-1}(\bar{\boldsymbol{\theta}}_T))(\tilde{L}_{\boldsymbol{\theta},T}(\boldsymbol{\theta}_0) - L_{\boldsymbol{\theta},T}(\boldsymbol{\theta}_0)) \\ &- L_{\boldsymbol{\theta\theta},T}^{-1}(\bar{\boldsymbol{\theta}}_T)(\tilde{L}_{\boldsymbol{\theta},T}(\boldsymbol{\theta}_0) - L_{\boldsymbol{\theta},T}(\boldsymbol{\theta}_0)) - L_{\boldsymbol{\theta},T}(\boldsymbol{\theta}_0)(\tilde{L}_{\boldsymbol{\theta\theta},T}^{-1}(\boldsymbol{\theta}_0) - L_{\boldsymbol{\theta\theta},T}^{-1}(\boldsymbol{\theta}_0)), \end{aligned}$$

and $\bar{\boldsymbol{\theta}}_T$ is a vector with elements $\bar{\boldsymbol{\theta}}_{j,t}$, $j = 1, \ldots, P+Q+1+d$ and for some $\alpha_j \in (0,1)$, $\bar{\boldsymbol{\theta}}_{j,T} = \alpha_j \boldsymbol{\theta}_{0,j} + (1-\alpha_j)\tilde{\boldsymbol{\theta}}_{j,T}$. Then, $T^{1/2}R_{1,T} \xrightarrow{a.s.} 0$ as $T \to \infty$ by Proposition 1, continuity of $L_{\boldsymbol{\theta\theta},T}(\cdot)$ and by Lemma C4 in the Appendix. From

(2.6) and Proposition 1 we have

$$T^{1/2}\begin{pmatrix}\tilde{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0\\ \gamma_{ac}\end{pmatrix}\overset{a.s.}{\sim}\begin{pmatrix}\mathcal{I}^{-1} & \mathbf{0}_{(P+Q+1+d)\times m}\\ \mathbf{0}_{m\times(P+Q+1+d)} & \boldsymbol{I}_{m\times m}\end{pmatrix}T^{1/2}\begin{pmatrix}L_{\boldsymbol{\theta},T}(\boldsymbol{\theta}_0)\\ \gamma_{ac}\end{pmatrix},$$

and the asymptotic normality follows by applying a central limit theorem for mixingales (see Proposition 2 in Scott, 1973) to $T^{1/2}(L_{\boldsymbol{\theta},T}(\boldsymbol{\theta}_0),\gamma_{ac})$, shown in Proposition 2 in the Appendix. The asymptotic covariance matrix of this vector is

$$\lim_{T\to\infty}\mathrm{Cov}\left(T^{1/2}\begin{pmatrix}L_{\boldsymbol{\theta},T}(\boldsymbol{\theta}_0)\\ \gamma_{ac}\end{pmatrix}\right)=\begin{pmatrix}\mathcal{I} & \boldsymbol{\Sigma}'_{l,\gamma_{ac}}\\ \boldsymbol{\Sigma}_{l,\gamma_{ac}} & \boldsymbol{\Sigma}_{\gamma_{ac}}\end{pmatrix},$$

from which the result follows. □

The next lemma establishes the asymptotic normality of the feasible autocorrelation function of $\tilde{u}_t(\tilde{\boldsymbol{\theta}}_T)$. The key is to approximate the feasible quantity $\tilde{\gamma}_{ac}$ as a linear combination of the unfeasible quantity γ_{ac} and the estimated parameters, and then use the previous Lemma 1 to establish the limiting normal distribution.

Lemma 2. *Under H_{ac}, for $P,Q>0$, $T^{1/2}\tilde{\boldsymbol{\rho}}_{ac}\xrightarrow{d}N(0,\boldsymbol{\Omega}_{ac})$.*

Proof. Linear approximation of the feasible autocovariance function around the true parameter value $\boldsymbol{\theta}_0$ gives

$$\begin{aligned}\tilde{\gamma}_{ac}=&\tilde{\gamma}_{ac}(\boldsymbol{\theta}_0)+\frac{\partial}{\partial\boldsymbol{\theta}'}\tilde{\gamma}_{ac}(\bar{\boldsymbol{\theta}}_T)(\tilde{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0)\\ =&\gamma_{ac}+\frac{\partial}{\partial\boldsymbol{\theta}'}\gamma_{ac}(\bar{\boldsymbol{\theta}}_T)(\tilde{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0)+R_{2,T},\end{aligned}$$

where

$$R_{2,T}=\tilde{\gamma}_{ac}(\boldsymbol{\theta}_0)-\gamma_{ac}+\left(\frac{\partial}{\partial\boldsymbol{\theta}'}\tilde{\gamma}_{ac}(\bar{\boldsymbol{\theta}})-\frac{\partial}{\partial\boldsymbol{\theta}'}\gamma_{ac}(\bar{\boldsymbol{\theta}})\right)(\tilde{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0),$$

and $T^{1/2}R_{2,T}\xrightarrow{a.s.}\mathbf{0}_{m\times 1}$ as $T\to\infty$ by Lemma D7 in the Appendix and by continuity of the derivatives. In Lemma D5 in the Appendix we justify that $\frac{\partial\boldsymbol{\theta}}{\partial\boldsymbol{\theta}'}\gamma_{ac}(\bar{\boldsymbol{\theta}}_T)\xrightarrow{a.s.}-\boldsymbol{\Sigma}_{l,\gamma_{ac}}$ as $T\to\infty$. Hence, by Lemma 1, continuity of $\frac{\partial}{\partial\boldsymbol{\theta}'}\gamma_{ac}(\cdot)$, Propositions 1, and the discussion above, $\tilde{\gamma}_{ac}$ has an asymptotic normal distribution with covariance matrix

$$\lim_{T\to\infty} \text{Cov}(T^{1/2}\tilde{\gamma}_{ac}) = \boldsymbol{\Sigma}_{l,\gamma_{ac}}\boldsymbol{\mathcal{I}}^{-1}\boldsymbol{\Sigma}'_{l,\gamma_{ac}} - \boldsymbol{\Sigma}_{l,\gamma_{ac}}\boldsymbol{\mathcal{I}}^{-1}\boldsymbol{\Sigma}'_{l,\gamma_{ac}} - \boldsymbol{\Sigma}_{l,\gamma_{ac}}\boldsymbol{\mathcal{I}}^{-1}\boldsymbol{\Sigma}'_{l,\gamma_{ac}} + \boldsymbol{\Sigma}_{\gamma_{ac}}$$

$$= \boldsymbol{\Sigma}_{\gamma_{ac}} - \boldsymbol{\Sigma}_{l,\gamma_{ac}}\boldsymbol{\mathcal{I}}^{-1}\boldsymbol{\Sigma}'_{l,\gamma_{ac}}.$$

Finally, note that

$$T\left(\tilde{\rho}_{i,ac} - \frac{\tilde{\gamma}_{i,ac}}{\sigma_0^2}\right) = T^{1/2}\tilde{\gamma}_{i,ac}\frac{T^{1/2}(\sigma_0^2 - \tilde{\gamma}_{i,ac})}{\sigma_0^2\tilde{\gamma}_{0,ac}} = O_p(1),$$

so that $\tilde{\boldsymbol{\rho}}_{ac} = \sigma_0^{-2}\tilde{\boldsymbol{\gamma}}_{ac} + O_p(T^{-1})$ and the asymptotic covariance matrix for the feasible autocorrelation function is

$$\begin{aligned}\boldsymbol{\Omega}_{ac} = \lim_{T\to\infty}\text{Cov}(T^{1/2}\tilde{\boldsymbol{\rho}}_{ac}) =& \boldsymbol{I}_{m\times m} - \sigma_0^{-4}\boldsymbol{\Sigma}_{l,\gamma_{ac}}\boldsymbol{\mathcal{I}}^{-1}\boldsymbol{\Sigma}'_{l,\gamma_{ac}} \\ =& \boldsymbol{I}_{m\times m} - \boldsymbol{\Lambda}_m\boldsymbol{\mathcal{I}}^{11}\boldsymbol{\Lambda}'_m,\end{aligned}$$

where $\boldsymbol{\mathcal{I}}^{11}$ is the inverse of the first $((P+Q)\times(P+Q))$ block of $\boldsymbol{\mathcal{I}}$.

□

Proposition 1 suggests estimating $\boldsymbol{\mathcal{I}}^{-1}$ by the inverse of the feasible Hessian matrix, $-\tilde{L}_{\boldsymbol{\theta\theta},T}(\tilde{\boldsymbol{\theta}}_T)^{-1}$, which is obtained during the estimation routine. Matrix $\boldsymbol{\Sigma}_{l,\gamma_{ac}}$ can be consistently estimated by replacing the model parameters by their estimated counterparts. If this matrix is denoted by $\tilde{\boldsymbol{\Omega}}_{ac} = \boldsymbol{I}_{m\times m} - \tilde{\boldsymbol{\Lambda}}_m\tilde{\boldsymbol{\mathcal{I}}}^{11}\tilde{\boldsymbol{\Lambda}}'_m$, then the asymptotic χ^2_m distribution follows by standard arguments.

Theorem 1. *Under $H_{0,ac}$, for $P, Q > 0$,*

$$Q_{ac} \xrightarrow{d} \chi^2_m, \quad \textit{as } T\to\infty.$$

2.3.2 Limiting distribution of Q_{hs}

The form of Q_{hs} suggests that there is less adjustment needed for the standard McLeod-Li Q statistic than there was for the Box-Pierce Q statistic. In particular, as long as the feasible error terms $\tilde{u}_t(\tilde{\boldsymbol{\theta}}_T)$ are calculated recursively top-down, as suggested in the previous section, then $T^{1/2}\tilde{\boldsymbol{\rho}}_{hs}$ is asymptotically

$N(0, \boldsymbol{I}_{m\times m})$ distributed, and the standard χ^2_m limiting distribution applies for Q_{hs}.

Limiting distribution of $\tilde{\boldsymbol{\rho}}_{hs}$ is invariant to the estimation uncertainty; the limiting behavior of the parameter estimates do not affect the limiting distribution of the test statistic.

Let $\kappa_t = \varepsilon_t^2 - \sigma_0^2$, and $\boldsymbol{\Sigma}_{\gamma_{hs}} = \mathrm{E}[\kappa^2]^2 \boldsymbol{I}_{m\times m}$. This matrix is finite under Assumption 1 (a) in Appendix A.

Lemma 3. *Under $H_{0,hs}$, for $P, Q > 0$, $T^{1/2}(\tilde{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0, \gamma_{hs}) \xrightarrow{d} N(0, \boldsymbol{\Sigma}_{\tilde{\boldsymbol{\theta}},\gamma_{hs}})$, where*

$$\boldsymbol{\Sigma}_{\tilde{\boldsymbol{\theta}},\gamma_{hs}} = \begin{pmatrix} \mathcal{I}^{-1} & \mathbf{0}_{(P+Q+1+d)\times m} \\ \mathbf{0}_{m\times(P+Q+1+d)} & \boldsymbol{\Sigma}_{\gamma_{hs}} \end{pmatrix}.$$

Proof. Steps in the proof of Lemma 1 above give

$$T^{1/2}\begin{pmatrix} \tilde{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0 \\ \gamma_{hs} \end{pmatrix} \overset{as.}{\sim} \begin{pmatrix} \mathcal{I}^{-1} & \mathbf{0}_{(P+Q+1+d)\times m} \\ \mathbf{0}_{m\times(P+Q+1+d)} & \boldsymbol{I}_{m\times m} \end{pmatrix} T^{1/2}\begin{pmatrix} L_{\boldsymbol{\theta},T}(\boldsymbol{\theta}_0) \\ \gamma_{hs} \end{pmatrix}.$$

In Proposition 2 in the Appendix we establish the limiting normal distribution of $T^{1/2}(L_{\boldsymbol{\theta},T}(\boldsymbol{\theta}_0), \gamma_{hs})$, and that

$$\lim_{T\to\infty} \mathrm{Cov}\left(T^{1/2}\begin{pmatrix} L_{\boldsymbol{\theta},T}(\boldsymbol{\theta}_0) \\ \gamma_{hs} \end{pmatrix}\right) = \begin{pmatrix} \mathcal{I} & \mathbf{0}_{(P+Q+1+d)\times m} \\ \mathbf{0}_{m\times(P+Q+1+d)} & \boldsymbol{\Sigma}_{\gamma_{hs}} \end{pmatrix},$$

and the stated result follows. □

The following Lemma gives the limiting normal distribution of the autocorrelation function of the squared feasible quantities $\tilde{u}_t(\tilde{\boldsymbol{\theta}}_T)^2$.

Lemma 4. *Under $H_{0,hs}$, for $P, Q > 0$, $T^{1/2}\tilde{\boldsymbol{\rho}}_{hs} \xrightarrow{d} N(0, \boldsymbol{I}_{\boldsymbol{m}\times\boldsymbol{m}})$.*

Proof. Following the steps in the proof of Lemma 2, linear approximation of the feasible statistic $\tilde{\gamma}_{hs}$ around the true parameter value $\boldsymbol{\theta}_0$ gives

$$\tilde{\gamma}_{hs} = \gamma_{hs} + \frac{\partial}{\partial\boldsymbol{\theta}'}\gamma_{hs}(\bar{\boldsymbol{\theta}}_T)(\tilde{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) + R_{3,T},$$

where

$$R_{3,T} = \tilde{\gamma}_{hs}(\boldsymbol{\theta}_0) - \gamma_{hs} + \left(\frac{\partial}{\partial\boldsymbol{\theta}'}\tilde{\gamma}_{hs}(\bar{\boldsymbol{\theta}}) - \frac{\partial}{\partial\boldsymbol{\theta}'}\gamma_{hs}(\bar{\boldsymbol{\theta}})\right)(\tilde{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0),$$

and $T^{1/2}R_{3,T} \stackrel{a.s.}{\rightarrow} \mathbf{0}_{m\times 1}$ as $T \rightarrow \infty$ by Proposition 1 and Lemma D7 in the Appendix. In lemma D5 we establish that $\frac{\partial}{\partial \boldsymbol{\theta}'}\gamma_{hs} \stackrel{a.s.}{\rightarrow} \mathbf{0}_{m\times(P+Q+1+d)}$, thus

$$T^{1/2}\tilde{\gamma}_{hs} \stackrel{as.}{\sim} \begin{pmatrix}\mathbf{0}_{m\times(P+Q+1+d)} & \boldsymbol{I}_{m\times m}\end{pmatrix} T^{1/2}\begin{pmatrix}\tilde{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0 \\ \gamma_{hs}\end{pmatrix},$$

and $\lim_{T\rightarrow\infty} \mathrm{Cov}(T^{1/2}\tilde{\gamma}_{hs}) = \boldsymbol{\Sigma}_{\gamma_{hs}}$. Note that

$$T\left(\frac{\tilde{\gamma}_{i,hs}}{\tilde{\gamma}_{0,hs}} - \frac{\tilde{\gamma}_{i,hs}}{\mathrm{E}[\kappa_t^2]}\right) = T^{1/2}\tilde{\gamma}_{i,hs}\frac{T^{1/2}(\mathrm{E}[\kappa_t^2] - \tilde{\gamma}_{i,hs})}{\mathrm{E}[\kappa_t^2]\tilde{\gamma}_{0,hs}} = O_p(1),$$

thus $T^{1/2}\tilde{\boldsymbol{\rho}}_{hs} = T^{1/2}\mathrm{E}[\kappa_t^2]^{-1}\tilde{\gamma}_{hs} + O_p(T^{-1/2})$, and $\lim_{T\rightarrow\infty}\mathrm{Cov}(T^{1/2}\tilde{\boldsymbol{\rho}}_{hs}) = \boldsymbol{I}_{m\times m}$. □

The χ_m^2-distribution follows in the standard manner for the test statistic Q_{hs}.

Theorem 2. *Under $H_{0,hs}$, for $P, Q > 0$,*

$$Q_{hs} \stackrel{d}{\rightarrow} \chi_m^2 \quad \text{as } T \rightarrow \infty.$$

2.3.3 Further consideration of the test statistics

In subsections 2.3.1 and 2.3.2 we developed the asymptotic theory for the test statistics and showed that the limit is a χ^2 distribution. Another way to characterize the distribution of $\tilde{\boldsymbol{\rho}}_{ac}$ is to note that, for an $(m \times 1)$ normally distributed random variable $\boldsymbol{z}$, $\boldsymbol{z}'\boldsymbol{z} \sim \sum_{i=1}^m \xi_i Z_i^2$, where $Z_i \sim N(0,1)$ and ξ_i is the i^{th} eigenvalue of the covariance matrix $\mathrm{Cov}(\boldsymbol{z})$. For Q_{hs}, the asymptotic χ_m^2 distribution follows by noting that the asymptotic covariance matrix of $\tilde{\boldsymbol{\rho}}_{hs}$, $\boldsymbol{I}_{m\times m}$, has m eigenvalues equal to one.

Next we provide some insight on how our theory overlaps with the findings of Box and Pierce (1970). To this end, it is useful to consider a particular partition of the matrix $\boldsymbol{\Lambda}_m$,

$$\boldsymbol{\Lambda}_m = \begin{pmatrix} \psi_{0,0}^{(a)} & \cdots & \psi_{0,1-P}^{(a)} & -\psi_{0,0}^{(b)} & \cdots & -\psi_{0,1-Q}^{(b)} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \psi_{0,m-1}^{(a)} & \cdots & \psi_{0,m-P}^{(a)} & -\psi_{0,m-1}^{(b)} & \cdots & -\psi_{0,m-Q}^{(b)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\zeta}_1^{(a)} & \boldsymbol{\zeta}_1^{(b)} \\ \vdots & \vdots \\ \boldsymbol{\zeta}_m^{(a)} & \boldsymbol{\zeta}_m^{(b)} \end{pmatrix}$$

$$= \begin{pmatrix} \boldsymbol{\Lambda}_m^{(a)} & \boldsymbol{\Lambda}_m^{(b)} \end{pmatrix}.$$

Let $\boldsymbol{\Lambda}_\infty^{(a)} = \lim_{m\to\infty} \sum_{i=1}^m \boldsymbol{\zeta}_i^{(a)\prime} \boldsymbol{\zeta}_i^{(a)}$, $\boldsymbol{\Lambda}_\infty^{(ab)} = \lim_{m\to\infty} \sum_{i=1}^m \boldsymbol{\zeta}_i^{(a)\prime} \boldsymbol{\zeta}_i^{(b)}$ and $\boldsymbol{\Lambda}_\infty^{(b)} = \lim_{m\to\infty} \sum_{i=1}^m \boldsymbol{\zeta}_i^{(b)\prime} \boldsymbol{\zeta}_i^{(b)}$. An exact form for the block diagonal matrix $\boldsymbol{\mathcal{I}}$ is given in Appendix C. If $\boldsymbol{\mathcal{I}}_{11}$ denotes the first $(P+Q)\times(P+Q)$ diagonal block, we have

$$\boldsymbol{\mathcal{I}}_{11} = \begin{pmatrix} \mathrm{E}[e_{x,t}^2]\boldsymbol{\Lambda}_\infty^{(a)} & \boldsymbol{\Lambda}_\infty^{(ab)} \\ \boldsymbol{\Lambda}_\infty^{(ab)\prime} & \mathrm{E}[e_{x,t}^2]\boldsymbol{\Lambda}_\infty^{(b)} \end{pmatrix},$$

where $e_{x,t} = \frac{\partial}{\partial \boldsymbol{\theta}} f_\eta(\varepsilon_t;\boldsymbol{\lambda}_0)/f_\eta(\varepsilon_t;\boldsymbol{\lambda}_0)$ (see Appendix 2.A). In the case of purely causal AR(P) model, matrix $\boldsymbol{\Lambda}_m$ takes the form $\boldsymbol{\Lambda}_m^{(a)}$, and $\boldsymbol{\mathcal{I}}_{11}$ is $\mathrm{E}[e_{x,t}^2]\boldsymbol{\Lambda}_\infty^{(a)}$. Now, the asymptotic covariance matrix of $T^{1/2}\tilde{\boldsymbol{\rho}}_{ac}$ in Lemma 2 is $\boldsymbol{I}_{m\times m} - \mathrm{E}[e_{x,t}^2]^{-1}\boldsymbol{\Lambda}_m^{(a)}(\boldsymbol{\Lambda}_\infty^{(a)})^{-1}\boldsymbol{\Lambda}_m^{(a)\prime}$. Following the idea of Box and Pierce (1970), for sufficiently large m, $\boldsymbol{\Lambda}_m^{(a)\prime}\boldsymbol{\Lambda}_m^{(a)} \approx \boldsymbol{\Lambda}_\infty^{(a)}$, so the matrix $\boldsymbol{\Lambda}_m^{(a)}(\boldsymbol{\Lambda}_\infty^{(a)})^{-1}\boldsymbol{\Lambda}_m^{(a)\prime}$ is approximately a projection matrix with P eigenvalues equal to zero and $m-P$ equal to one. In Gaussian case $\mathrm{E}[e_{x,t}^2] = 1$, and the matrix $\boldsymbol{I}_{m\times m} - \mathrm{E}[e_{x,t}^2]^{-1}\boldsymbol{\Lambda}_m^{(a)}(\boldsymbol{\Lambda}_\infty^{(a)})^{-1}\boldsymbol{\Lambda}_m^{(a)\prime}$ would be approximately a projection with $m-P$ eigenvalues of one, and the usual χ^2_{m-P} approximation would hold. In the non-Gaussian case $\mathrm{E}[e_{x,t}^2] \neq 1$ (Andrews et al., 2006, Remark 2), and we see how this term distorts the approximation even in the large samples. However, our modified test Q_{ac} will correct for this bias, at the cost of having to estimate the matrix $\boldsymbol{\mathcal{I}}$. It is worth noting, and easy to verify, that the purely noninvertible MA(Q) case is analogous. In the Gaussian case χ^2_{m-Q} approximation holds, although our modified test Q_{ac} accounts for the distortion of non-Gaussian distribution as well.

These ideas generalize to the noninvertible ARMA(P,Q) in (2.1) as well. The asymptotic covariance of $T^{1/2}\tilde{\boldsymbol{\rho}}_{ac}$ is given in Lemma 2. Using approximation $\boldsymbol{\Lambda}_m'\boldsymbol{\Lambda}_m \approx \boldsymbol{\Lambda}_\infty$, we have $\boldsymbol{\Lambda}_m\boldsymbol{\mathcal{I}}^{11}\boldsymbol{\Lambda}_m'\boldsymbol{\Lambda}_m\boldsymbol{\mathcal{I}}^{11}\boldsymbol{\Lambda}_m' \approx \boldsymbol{\Lambda}_m\boldsymbol{\mathcal{I}}^{11}\boldsymbol{\Lambda}_\infty\boldsymbol{\mathcal{I}}^{11}\boldsymbol{\Lambda}_m'$, so we can conclude that the asymptotic covariance matrix is not a projection, thus the χ^2_{m-P-Q} approximation fails. Although the Gaussianity of the error term was ruled out in order to achieve the identification of the model, a point that we will clarify below in an example, it is illustrative to think about what happens when $\mathrm{E}[e_{x,t}^2] = 1$. In this case $\boldsymbol{\mathcal{I}}^{11}\boldsymbol{\Lambda}_\infty = \boldsymbol{I}_{(P+Q)\times(P+Q)}$, and the the covariance matrix would be a projection matrix. In general, non-Gaussianity breaks down this property.

2.3.4 Example: noninvertible ARMA(1,1)

Let us illustrate the method and consider the simplest possible noninvertible ARMA(1,1) case

$$(1 - a_{0,1}B)y_t = (1 - b_{0,1}B^{-1})\varepsilon_t.$$

Matrices in the asymptotic covariance matrix $\mathbf{\Omega}_{ac}$ are

$$\mathbf{\Lambda}_m = \begin{pmatrix} 1 & -1 \\ a_{0,1} & -b_{0,1} \\ \vdots & \vdots \\ a_{0,1}^{m-1} & -b_{0,1}^{m-1} \end{pmatrix} \text{ and } \mathcal{I}_{11} = \begin{pmatrix} \mathrm{E}[e_{x,t}^2](1-a_{0,1}^2)^{-1} & -(1-a_{0,1}b_{0,1})^{-1} \\ -(1-a_{0,1}b_{0,1})^{-1} & \mathrm{E}[e_{x,t}^2](1-b_{0,1}^2)^{-1} \end{pmatrix}.$$

As pointed out in Lanne and Saikkonen (2013), the vitality of the non-Gaussianity assumption can be seen in matrix $\mathcal{I}_{11}$ above. If $\mathrm{E}[e_{x,t}^2] = 1$, the matrix would be singular for $a_{0,1} = b_{0,1}$. More generally, this singularity occurs whenever the roots of $a_0(z)$ are reciprocal to the roots of $b_0(z)$. Non-Gaussianity ensures nonsingularity of $\mathcal{I}_{11}$ even if the true DGP is weak white noise, i.e. $a_{0,1} = b_{0,1}$.

2.4 Monte Carlo simulations

2.4.1 Size simulations

In this section we study the finite sample properties of the proposed test statistics using Monte Carlo simulations. We begin with size simulations using two different data generating processes compatible with model (2.1):

$$\begin{aligned} \text{DGP I:} \quad & y_t = 0.2y_{t-1} + \varepsilon_t - 0.2\varepsilon_{t+1}, \quad \text{and} \\ \text{DGP II:} \quad & y_t = 0.2y_{t-1} + \varepsilon_t - 0.8\varepsilon_{t+1}. \end{aligned}$$

Throughout this exercise we set $\sigma_0^2 = 2$ and $m = 5$. The error process is $\varepsilon_t = \sigma_0 \eta_t$ and η_t follows the Student's t-distribution with degrees of freedom λ_0. We vary the sample size as $T = 250; 500$ and 10,000. The smallest sample size $T = 250$ represents a magnitude often encountered when, for example, quarterly economic data is used. This is the case in our empirical example in the next section, where we use quarterly stock return data from 1947Q1 to 2007Q4. $T = 500$ is also relevant for lower frequency financial and macro economic data, for example when monthly data is under consideration for shorter periods of time. The largest sample size is used to illustrate the asymptotic properties of the statistics.

We use three different degrees of freedom, $\lambda_0 = 3, 5$ and 9, for the Student's t-distribution. Assumption 1 in Section 2.A in Appendix lays down the moment conditions of the error term process ε_t: asymptotic properties of the Q_{ac} and Q_{hs} tests have been derived under the assumptions of finite fourth and eight moments, respectively. These assumptions are satisfied for $\lambda_0 > 4$ for the Q_{ac} statistic and $\lambda_0 > 8$ for the Q_{hs} statistic. Our selection of the degrees of freedom parameter allows us to study the properties of the tests when the assumptions are met, but also illustrate the consequences of the deviations from these conditions. For $\lambda_0 = 3$, moment conditions fail to hold for both of the tests, and the deviation from this assumption is more severe for the Q_{hs} statistic. If $\lambda_0 = 5$, the conditions of Q_{ac} are satisfied, but those of Q_{hs} are not. When $\lambda_0 = 9$, assumptions are satisfied for both of the statistics.

For each combinations of T and λ_0, we simulate 1,000 data sets using DGP I and DGP II. To avoid initialization effects, 2,000 extra observations at the beginning of each series are simulated and discarded. We fit the noninvertible ARMA(1,1) model to each of the series and use the residuals to perform the tests. Summaries of the simulation results are presented graphically.[9] Figure 2.1 plots the discrepancy of Q_{ac} test size: the deviation of the test's actual size from its nominal size is plotted against the nominal size for significance levels 1%, 1.1%,...,10%. Columns in Figure 2.1 refer to different values of λ_0. For all the parameter combinations, the size properties are adequate for large samples. The moment conditions, however, play an important role for the

[9] Additional simulation results are available upon request. The size and power properties have been investigated using different parameter value combinations, and standard Ljung-Box tests have also been calculated for the sake of comparison. Not surprisingly, L-B test tends to overreject because it lacks the covariance matrix that offsets for the non-Gaussianity.

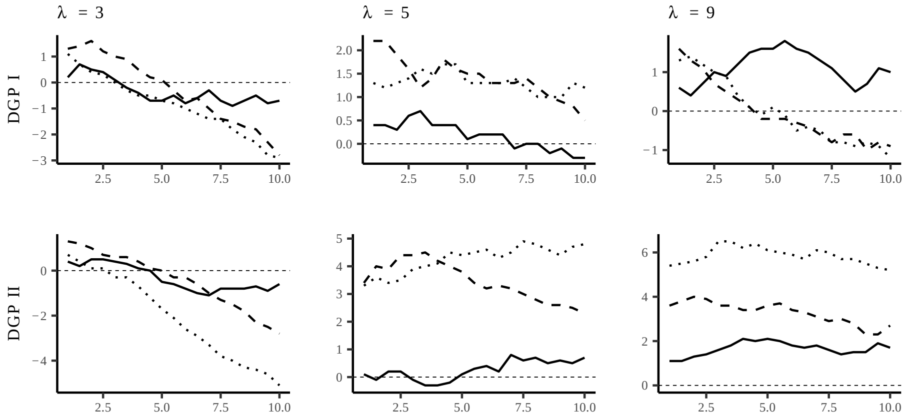

Figure 2.1: Size of the Q_{ac} test statistic with $m = 5$ for different sample sizes. The deviation of the test size from its nominal value is plotted against the nominal values and the deviation is measured in percentage points. Different columns correspond to different values of λ_0, and rows to different data generating processes, DGP I and DGP II, respectively. Solid line depicts the results for $T = 10,000$, dotted line for $T = 500$, and dashed line for $T = 250$.

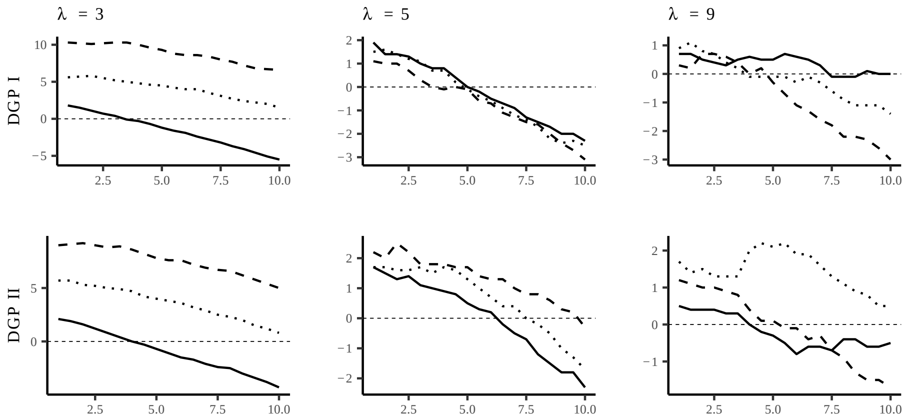

Figure 2.2: Size of the Q_{hs} test statistic with $m = 5$ for different sample sizes. The deviation of the test size from its nominal value is plotted against the nominal values and the deviation is measured in percentage points. Different columns correspond to different values of λ_0, and rows to different data generating processes, DGP I and DGP II, respectively. Solid line depicts the results for $T = 10,000$, dotted line for $T = 500$, and dashed line for $T = 250$.

smaller samples. When the moment conditions are not satisfied ($\lambda_0 = 3$, in the leftmost column), we find a downward-sloping pattern for the sizes. When $\lambda_0 = 9$ (the rightmost column), the distribution of the error process is close to Gaussian. We would expect to find difficulties in the estimation procedure. This might explain the overrejections we encounter in this case. The large sample size ensures that the true DGP is identified during the estimation, but for the smaller samples the estimated parameters might take false values and thus impose autocorrelation in the residuals. It seems to be the case that the higher the autocorrelation in the data (lower row), the higher the overrejection rate.

Q_{hs} test is more prone to moment conditions, as can be seen in Figure 2.2. When the assumptions are satisfied, the size of Q_{hs} is well aligned on its nominal values over the set of nominal values (the rightmost column). For the smaller sample sizes, the patterns are similar to those of the Q_{ac} test. The importance of the moment condition can be seen in the downward-sloping lines in the left and middle columns. The χ^2 distribution does not seem to capture the behavior of the test statistics. For the large samples, the size properties are tolerable even if the moment condition does not hold, and even for the smaller samples, the distortion is not that intolerable for $\lambda_0 = 5$.

2.4.2 Power simulations

Power properties of the Q_{ac} and Q_{hs} tests are studied by simulating data using three different models that are more general than the noninvertible ARMA(1,1): noninvertible ARMA(1,2) and two noninvertible ARMA(1,1) -ARCH(1) models with different magnitudes of heteroskedasticity. The model equations are

$$\begin{aligned}
&\text{DGP III:} \quad y_t = 0.2y_{t-1} + \varepsilon_t - 0.2\varepsilon_{t+1} - 0.2\varepsilon_{t+2}, \\
&\text{DGP IV:} \quad y_t = 0.2y_{t-1} + \sigma_t\eta_t - 0.2\sigma_{t+1}\eta_{t+1}, \quad \sigma_t = \sqrt{2 + 0.2\eta_{t-1}^2}, \quad \text{and} \\
&\text{DGP V:} \quad y_t = 0.2y_{t-1} + \sigma_t\eta_t - 0.2\sigma_{t+1}\eta_{t+1}, \quad \sigma_t = \sqrt{2 + 0.8\eta_{t-1}^2}.
\end{aligned}$$

DGP III is used for studying how well the test statistics can detect misspecified lead length in model (2.1). DGP IV and DGP V are noninvertible ARMA(1,1) models with ARCH-type heteroskedasticity, the processes studied in Meitz and Saikkonen (2013). These models are chosen to illustrate the power of Q_{hs} test against nonlinear processes with heteroskedasticity.

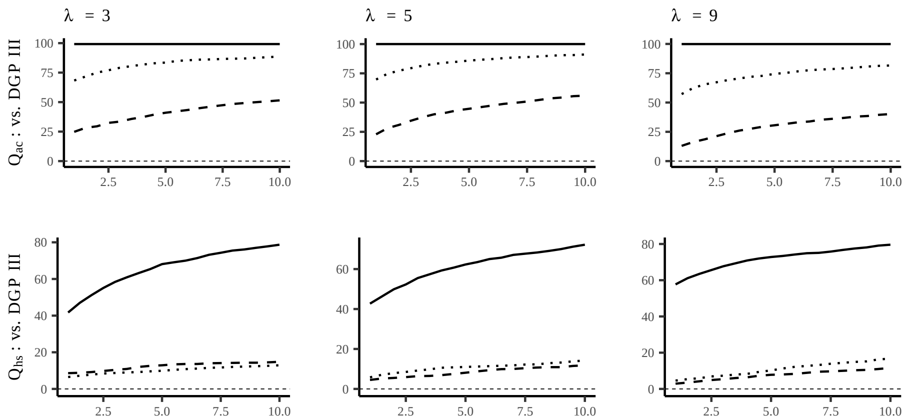

Figure 2.3: Power of the Q_{ac} and Q_{hs} test statistics against DGP III, for different sample sizes. The power of the tests are plotted against the significance levels, and power is measured in percentage points. Different columns correspond to different values of λ_0, and different rows correspond to different test statistics, Q_{ac} and Q_{hs}, respectively. Solid line depicts the results for $T = 10,000$, dotted line for $T = 500$, and dashed line for $T = 250$.

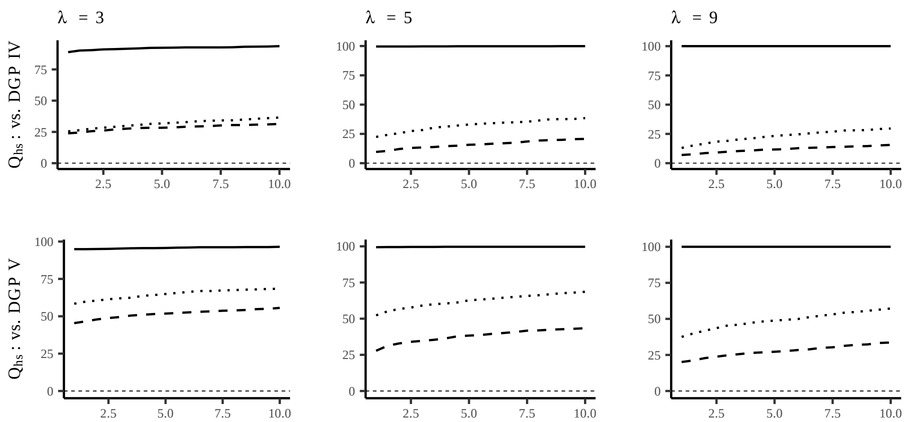

Figure 2.4: Power of the Q_{hs} test statistic against DGP IV (upper row) and DGP V (lower row), for different sample sizes. The power of the test is plotted against the significance levels, and power is measured in percentage points. Different columns correspond to different values of λ_0. Solid line depicts the results for $T = 10,000$, dotted line for $T = 500$, and dashed line for $T = 250$.

The design of the Monte Carlo experiment is similar to the size simulations. Again, we have simulated 1,000 data sets using the models described above, and we have fitted the noninvertible ARMA(1,1) model for each data set. The test statistics are calculated using the fitted residuals.

The test statistics' power to detect misspecified lag length is illustrated in Figure 2.3. Q_{ac} test statistic detects the autocorrelation induced by the misspecified MA-polynomial lead length at a high rate. Although the autocorrelation is assumable modest ($b_{0,2} = 0.2$), from the large samples, the misspecification is always detected. For more modest sample sizes, for 5% significance level, the power is around 75%, and never below 50% for any significance levels.

Heteroskedasticity implied by the noninvertible ARMA(1,1) model is usually very modest, so we would not expect the misspecified lag length to imply high heteroskedasticity for the residuals either. Nevertheless, the Q_{hs} statistic always has more than 40% power with even the smallest considered significance level, when the sample size is $T = 10,000$. With 5% significance level, the power is always more than 50%, and the power increases substantially when the sample size is doubled from 250 to 500. The sample size affects the power properties of Q_{hs} dramatically. For the smaller sample sizes, the misspecification is rarely detected via heteroskedasticity, but in the large samples, this modest heteroskedasticity can be used to identify the misspecified model quite accurately.

Even the modest ARCH-type heteroskedasticity can be detected with Q_{hs} very accurately in the large samples, as we can see in Figure 2.4. The power of the test is, however, very much dependent on the sample size. Increasing sample size from 250 to 500 makes a large difference in the power properties. The more severe heteroskedasticity in DGP V is detected more accurately even in the samples of more modest size. Depending on λ_0 and the significance level of the test, the power tends to be close to 50% or over, for $T = 500$, and over 25% for $T = 250$.

2.5 Empirical application

The question we address in this section is whetherthe diagnostic checks shed light on the predictability of asset returns. In our context, predictability simply means non-constant conditional expectation. According to dynamic asset pricing literature, predictability is a consequence of agents' risk aversion. For a thorough discussion, see Chapter 9 in Campbell, Lo, and MacKinlay (1997), or Chapter 2 in Singleton (2009). By no means should the predictability be manifested in a form of autocorrelation; rather we are expecting to encounter nonlinear predictability.

The advantage the noninvertible model has over the invertible one in modeling asset returns is its generality. In the previous analysis of predictability, testing has usually been based on the invertible and autocorrelated ARMA model, which is implied, for example, by the price-trend model of Taylor (1982) or the mean-reversion model of Poterba and Summers (1988). The noninvertible ARMA model is capable of capturing all the same autocorrelation structures as the invertible model, but is also capable of controlling for the nonlinearities often encountered in the financial time series data. As for the invertible ARMA model, the lack of autocorrelation automatically implies the independence of the data: for the noninvertible ARMA model, zero-autocorrelation is just a special case and the observations may still be dependent in a nonlinear fashion. This generality allows us to model a richer class of dependencies with noninvertible model, than a conventional invertible ARMA model would allow.[10]

Following Lanne et al. (2013), we suggest that the noninvertible ARMA(1,1) model is a particularly promising candidate to capture this nonlinear predictability. To our knowledge, this is the first time the model has been investigated from the standpoint of model fit based on asymptotic results. Preceding related work has mainly illustrated how the noninvertible model can mimic the nonlinear behavior of stock markets (Breidt et al., 2001), or how the predictability can be tested under the null of noninvertible ARMA model (Lanne et al., 2013). The evaluation of the model fit has been performed so far merely by looking at the sample autocorrelation functions, without having the correct critical values.

Using statistical tests on the estimated parameters of the noninvertible ARMA(1,1) model, Lanne et al. (2013) reported nonlinear predictability, in line with the asset pricing theory. Their testing procedure implicitly assumed that, under the null, the correct model is the noninvertible ARMA. We take another look at this data and show that our diagnostic checks actually support this assumption and thus support their conclusions of nonlinear predictability.

[10] An interesting special case of the noninvertible ARMA model (2.1) is the so-called all-pass model: $a_0(B)y_t = a_0(B^{-1})\varepsilon_t$. Whenever the roots of the AR polynomial coincide with the reciprocals of the roots of the MA polynomial, the process y_t is weak white noise. This can be seen by noting that the spectral density of the process y_t is constant $\sigma_0^2/(2\pi)$. Polynomials $a_0(B)$ and $a_0(B^{-1})$ do not cancel out, and the data is not iid. For example, the squared observations of the data can be shown to be correlated (Lanne et al., 2013, Appendix A.2. in).

In this section we apply our test statistics to evaluate the fit of the noninvertible ARMA models to the quarterly measured stock portfolio returns compiled of U.S. stocks. We use three value-weighted, size-ordered stock portfolios, and the market portfolio, which include data from NYSE, AMEX, and NASDAQ stocks from January 1947 to December 2007, the same data that was used by Lanne et al. (2013). Data is obtained from Kenneth French's website.[11] Monthly returns are transformed into quarterly quantities by continuous compounding and means are subtracted from the series.

Estimation results are presented in Table 2.1. The left side of the table shows the estimated values of the parameters. It is worth noting that the parameters are estimated with good precision, they are statistically different from zero, and AR and MA parameters are close to each other. This suggests that the series are very mildly autocorrelated, but dependent in some nonlinear way. Estimation has been based on the Student's t-distribution. The estimates of the degrees of freedom parameter λ_0 suggest that the innovation processes in all of the cases have finite fifth moments. This is enough to satisfy the moment condition imposed on the Q_{ac} test, but it fails to meet the assumption of the finite eight moment of the Q_{hs} test. Nevertheless, the Monte Carlo experiment in the previous section encourages us to still carry out the tests, with caution, as the size properties of the test were not too distorted by this relatively modest deviation from the moment condition.

The columns on the right give the p-values of the Q_{ac} and Q_{hs} tests for three different lag lengths, $m = 5$, 9, and 12. For three out of four portfolios, the null can not be rejected, suggesting that there is no autocorrelation left in the residuals or squared residuals. The noninvertible ARMA(1,1) model seems like an adequate model for the Market, Middle 40% and Top 30% portfolios in the light of our checks. The heteroskedasticity in the residuals of the noninvertible ARMA(1,1) model for the Bottom 30% portfolio can not be ruled out, but it turns out that the noninvertible ARMA(2,2) model is suitable to control for that.[12]

[11] `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`, downloaded Jan. 10, 2017

[12] All the estimated parameters of the noninvertible ARMA(2,2) model are statistically highly significant (all p-values $< .01$) and the p-values of the Q_{ac} and Q_{hs} tests with $m = 5$ are 0.783 and 0.143 for the Q_{ac} and Q_{hs} tests, respectively, and similar for different choices of m as well.

					$Q_{ac,T}$			$Q_{hs,T}$		
Portfolio	a	b	σ^2	λ	5	9	12	5	9	12
Market	.748 (.083)	.759 (.090)	8.074 (.673)	5.012 (1.803)	.982	.911	.961	.606	.845	.794
Bottom 30%	.846 (.039)	.936 (.037)	11.855 (.969)	5.285 (2.519)	1.000	.959	.942	.001	.001	.000
Middle 40%	.684 (.093)	.780 (.092)	9.826 (.751)	5.404 (2.117)	.880	.861	.924	.304	.376	.577
Top 30%	.746 (.081)	.721 (.092)	7.679 (.603)	5.152 (1.842)	1.000	.931	.963	.501	.850	.856

Table 2.1: The noninvertible ARMA(1,1) model has been estimated to four stock return index series. Table indicates the parameter estimates and their standard errors. Test statistics Q_{ac} and Q_{hs} have been calculated from the residuals of the fitted models and their p-values have been reported for three different lag lengths m for each test.

2.6 Conclusions

In this article we derived asymptotic properties for two residual-based test statistics for evaluating model adequacy of the noninvertible ARMA model. The Q_{ac} test statistic is designed to detect remaining autocorrelation in the residuals and it is analogous to the Box-Pierce Q-statistic in the standard causal and invertible case. The asymptotic distribution of the Q_{ac} test statistic is not invariant to the estimation uncertainty of the model, so it must be taken into account in the construction of the test. Also, the need to assume a non-Gaussian error process implies a more involved form for the test statistic than in the Box-Pierce setting. The Q_{hs} test statistic is designed to capture autocorrelation in the squared residuals, and detect possible heteroskedasticity in the residuals. This test, in turn, is invariant to the estimation uncertainty, so we found that the McLeod-Li Q-statistic is an asymptotically valid test for this purpose among the noninvertible models as well, as long as the noninvertibility is correctly taken into account in estimation of the model, and the residuals are calculated correctly recursively from the last to the first. Both tests have an asymptotic χ^2 distribution. We also showed that the test statistics have adequate size properties and power against different types of misspecifications.

Our empirical example was designed to evaluate the adequacy of the noninvertible ARMA model to the quarterly U.S. stock return data. The model was found, in light of our tests, a potential candidate in modeling these mildly autocorrelated, but possibly nonlinearly dependent, data. This finding provides good grounds for looking for nonlinear predictability in the asset returns. Work in this direction has recently been done by Lanne et al. (2013), where the noninvertibility was implicitly assumed in their testing procedure. Our findings thus support their assumption and moreover their conclusions of possible nonlinear predictability.

In this article we based the asymptotic properties of the tests on ML estimators of the model parameters. We do note that there are other possible estimation methods available as well, for example the least absolute deviation method by Breidt et al. (2001) and Wu and Davis (2010). These methods may lack some of the efficiency of the ML method, but there are certain benefits to not having to define the error distribution. We would expect to find that, with modest modifications, our diagnostic testing strategy could incorporate other estimation methods as well, as long as the non-Gaussianity and some moment conditions hold for the error process.

References

Alessi, L., M. Barigozzi, and M. Capasso (2011). Non-fundamentalness in structural econometric models: A review. *International Statistical Review 79*, 16–47.

Andrews, B., M. Calder, and R. A. Davis (2009). Maximum likelihood estimation for α-stable autoregressive processes. *Annals of Statistics 37*, 1946–1982.

Andrews, B., R. A. Davis, and F. J. Breidt (2006). Maximum likelihood estimation for all-pass time series models. *Journal of Multivariate Analysis 97*, 1638 – 1659.

Blanchard, O. and R. Perotti (2002). An empirical characterization of the dynamic effects of changes in government spending and taxes on output. *Quarterly Journal of Economics 117*, 1329–1368.

Box, G. E. and D. A. Pierce (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association 65*, 1509–1526.

Breidt, F. J., R. A. Davis, and A. A. Trindade (2001). Least absolute deviation estimation for all-pass time series models. *Annals of Statistics 29*, 919–946.

Campbell, J. Y., A. W.-C. Lo, and A. C. MacKinlay (1997). *The Econometrics of Financial Markets*, Volume 2. Princeton University Press, Princeton, NJ.

Cavaliere, G., H. B. Nielsen, and A. Rahbek (2018). Bootstrapping noncausal autoregressions: With applications to explosive bubble modeling. *Journal of Business & Economic Statistics*. forthcoming.

Chen, W. W. and R. S. Deo (2004). A generalized portmanteau goodness-of-fit test for time series models. *Econometric Theory 20*, 382–416.

Cui, Y., T. J. Fisher, and R. Wu (2014). Diagnostic tests for non-causal time series with infinite variance. *Journal of Statistical Planning and Inference 147*, 117–131.

Fernández-Villaverde, J., J. F. Rubio-Ramírez, T. J. Sargent, and M. W. Watson (2007). ABCs (and Ds) of understanding VARs. *American Economic Review 97*, 1021–1026.

Forni, M. and L. Gambetti (2014). Sufficient information in structural VARs. *Journal of Monetary Economics 66*, 124–136.

Francq, C., R. Roy, and J.-M. Zakoïan (2012). Diagnostic checking in arma models with uncorrelated errors. *Journal of the American Statistical Association*.

Friés, S. and J.-M. Zakoïan (2019). Mixed causal-noncausal ar processes and the modelling of explosive bubbles. *Econometric Theory*. Forthcoming.

Gouriéroux, C. and J.-M. Zakoïan (2017). Local explosion modelling by noncausal process. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 79*, 737–756.

Hansen, L. P. and T. J. Sargent (1981). Exact linear rational expectations models: Specification and estimation. Technical report, Federal Reserve Bank of Minneapolis.

Hansen, L. P. and T. J. Sargent (1991). Two difficulties in interpreting vector autoregressions. In L. P. Hansen, T. J. Sargent, J. Heaton, A. Marcet, and W. Roberds (Eds.), *Rational expectations econometrics*. Oxford: Westview Press Boulder, CO.

Hong, Y. (1999). Hypothesis testing in time series via the empirical characteristic function: a generalized spectral density approach. *Journal of the American Statistical Association 94*, 1201–1220.

Hong, Y. and T.-H. Lee (2003). Diagnostic checking for the adequacy of nonlinear time series models. *Econometric Theory 19*, 1065–1121.

Huang, J. and Y. Pawitan (2000). Quasi-likelihood estimation of non-invertible moving average processes. *Scandinavian Journal of Statistics 27*, 689–702.

Kasa, K., T. B. Walker, and C. H. Whiteman (2014). Heterogeneous beliefs and tests of present value models. *Review of Economic Studies 81*, 1137–1163.

Lanne, M., M. Meitz, and P. Saikkonen (2013). Testing for linear and nonlinear predictability of stock returns. *Journal of Financial Econometrics 11*, 682–705.

Lanne, M. and P. Saikkonen (2013). Noncausal vector autoregression. *Econometric Theory 29*, 447–481.

Leeper, E. M., T. B. Walker, and S.-C. S. Yang (2013). Fiscal foresight and information flows. *Econometrica 81*, 1115–1145.

Li, W. and A. McLeod (1988). ARMA modelling with non-Gaussian innovations. *Journal of Time Series Analysis 9*, 155–168.

Lii, K.-S. and M. Rosenblatt (1992). An approximate maximum likelihood estimation for non-gaussian non-minimum phase moving average processes. *Journal of Multivariate Analysis 43*, 272–299.

Lii, K.-S. and M. Rosenblatt (1996). Maximum likelihood estimation for non-Gaussian nonminimum phase ARMA sequences. *Statistica Sinica 6*, 1–22.

Lin, J.-W. and A. I. McLeod (2008). Portmanteau tests for ARMA models with infinite variance. *Journal of Time Series Analysis 29*, 600–617.

Ljung, G. M. and G. E. Box (1978). On a measure of lack of fit in time series models. *Biometrika 65*, 297–303.

Lobato, I., J. C. Nankervis, and N. E. Savin (2001). Testing for autocorrelation using a modified box-pierce Q test. *International Economic Review 42*, 187–205.

Lobato, I. N. (2001). Testing that a dependent process is uncorrelated. *Journal of the American Statistical Association 96*, 1066–1076.

Lobato, I. N., J. C. Nankervis, and N. Savin (2002). Testing for zero autocorrelation in the presence of statistical dependence. *Econometric Theory 18*, 730–743.

McLeod, A. I. and W. K. Li (1983). Diagnostic checking ARMA time series models using squared-residual autocorrelations. *Journal of Time Series Analysis 4*, 269–273.

Meitz, M. and P. Saikkonen (2011). Parameter estimation in nonlinear AR-GARCH models. *Econometric Theory 27*, 1236–1278.

Meitz, M. and P. Saikkonen (2013). Maximum likelihood estimation of a non-invertible ARMA model with autoregressive conditional heteroskedasticity. *Journal of Multivariate Analysis 114*, 227–255.

Monti, A. C. (1994). A proposal for a residual autocorrelation test in linear models. *Biometrika 81*, 776–780.

Peña, D. and J. Rodríguez (2002). A powerful portmanteau test of lack of fit for time series. *Journal of the American Statistical Association 97*, 601–610.

Poterba, J. M. and L. H. Summers (1988). Mean reversion in stock prices. *Journal of Financial Economics 22*, 27 – 59.

Romano, J. P. and L. A. Thombs (1996). Inference for autocorrelations under weak assumptions. *Journal of the American Statistical Association 91*, 590–600.

Rosenblatt, M. (2012). *Gaussian and non-Gaussian linear time series and random fields*. Springer Science & Business Media.

Scott, D. J. (1973). Central limit theorems for martingales and for processes with stationary increments using a Skorokhod representation approach. *Advances in Applied Probability 5*, 119–137.

Singleton, K. J. (2009). *Empirical dynamic asset pricing: model specification and econometric assessment*. Princeton University Press.

Straumann, D. and T. Mikosch (2006). Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equations approach. *Annals of Statistics 34*, 2449–2495.

Taylor, S. J. (1982). Tests of the random walk hypothesis against a price-trend hypothesis. *Journal of Financial and Quantitative Analysis 17*, 37–61.

Velasco, C. and I. N. Lobato (2018). Frequency domain minimum distance inference for possibly noninvertible and noncausal ARMA models. *Annals of Statistics 46*, 555–579.

Wu, R. and R. A. Davis (2010). Least absolute deviation estimation for general autoregressive moving average time-series models. *Journal of Time Series Analysis 31*, 98–112.

Appendix

2.A Assumptions

This section lists the assumptions under which the main results are derived in the main text. Let us introduce some notations. Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space and $\mathcal{F}_t$ a σ-algebra generated by random variables $\{\eta_s\}_{s<t}$. The true but unknown parameters of the model are $\boldsymbol{\theta}_{0,a} = (a_{0,1}, \ldots, a_{0,P})$, $\boldsymbol{\theta}_{0,b} = (b_{0,1}, \ldots, b_{0,Q})$, σ_0, and $\boldsymbol{\lambda}_0 = (\lambda_{0,1}, \ldots, \lambda_{0,d})$. Let us collect these parameters into a $(P+Q+1+d)$ vector $\boldsymbol{\theta}_0 = (\boldsymbol{\theta}_{0,a}, \boldsymbol{\theta}_{0,b}, \sigma_0, \boldsymbol{\lambda}_{\mathbf{0}})$.

Polynomials $a(z, \boldsymbol{\theta}) = 1 - a_1 z - \cdots - a_P z^P$ and $b(z^{-1}, \boldsymbol{\theta}) = 1 - b_1 z^{-1} - \cdots - b_Q z^{-Q}$ define a counterpart for the model (2.1) for any parameter values $\boldsymbol{\theta} = (\boldsymbol{\theta}_a, \boldsymbol{\theta}_b, \sigma, \boldsymbol{\lambda})$, with $\boldsymbol{\theta} = (a_1, \ldots, a_P)$ and $\boldsymbol{\theta}_b = (b_1, \ldots, b_Q)$. Clearly, in model (2.1), $a_0(z) = a(z, \boldsymbol{\theta}_0)$ and $b_0(z) = b(z, \boldsymbol{\theta}_0)$.

The partial derivative of the density function $f_\eta(x; \boldsymbol{\lambda})$ is denoted by sub-index: $f_{\eta,x}(x; \boldsymbol{\lambda}) = \frac{\partial}{\partial x} f_\eta(x; \boldsymbol{\lambda})$, $f_{\eta,\lambda}(x; \boldsymbol{\lambda}) = \frac{\partial}{\partial \lambda} f_\eta(x; \boldsymbol{\lambda})$, and $f_{\eta,yz}(x; \boldsymbol{\lambda}) = \frac{\partial}{\partial x \partial y'} f_\eta(x; \boldsymbol{\lambda})$ for $y, z \in \{x, \boldsymbol{\lambda}\}$. For the derivatives of the log-likelihood, we use the shorthand notations

$$e_{x,t}(\boldsymbol{\theta}) = \frac{f_{\eta,x}(\sigma^{-1} u_t(\boldsymbol{\theta}); \boldsymbol{\lambda})}{f_\eta(\sigma^{-1} u_t(\boldsymbol{\theta}); \boldsymbol{\lambda})} \quad \text{and} \quad e_{\boldsymbol{\lambda},t}(\boldsymbol{\theta}) = \frac{f_{\eta,\lambda}(\sigma^{-1} u_t(\boldsymbol{\theta}); \boldsymbol{\lambda})}{f_\eta(\sigma^{-1} u_t(\boldsymbol{\theta}); \boldsymbol{\lambda})}.$$

When evaluated at the true parameter value $\boldsymbol{\theta}_0$, we use the shorthand notation $e_{x,t}(\boldsymbol{\theta}_0) = e_{x,t}$.

The first assumption summarizes the restrictions imposed on the error process ε_t.

Assumption 1. *The error process is $\varepsilon_t = \sigma_0 \eta_t$ with η_t an iid sequence with $E[\eta_t] = 0$ and $E[\eta_t^2] = 1$. The distribution of η_t is symmetric and non-Gaussian with the density function $f_\eta(x; \boldsymbol{\lambda}_0)$. In addition to the finite second moments, the process has either*

(a) finite fourth moments, $E[\eta_t^4] < \infty$, or

(b) finite eight moments, $E[\eta_t^8] < \infty$.

Polynomials $a(z, \boldsymbol{\theta})$ and $b(z^{-1}; \boldsymbol{\theta})$ satisfy the causality and invertibility conditions for all $\boldsymbol{\theta} \in \Theta$, where Θ is the permissible parameter space.

Assumption 2. *The permissible parameter space is* $\Theta = \Theta_a \times \Theta_b \times \Theta_\sigma \times \Theta_{\boldsymbol{\lambda}}$*, where*

$$\begin{aligned}\Theta_a =& \{\boldsymbol{\theta}_a \in \mathbb{R}^P \; ; \; a(z) \neq 0 \; \forall \; |z| \leq 1\},\\ \Theta_b =& \{\boldsymbol{\theta}_b \in \mathbb{R}^Q \; ; \; b(z^{-1}) \neq 0 \; \forall \; |z^{-1}| \leq 1\},\\ \Theta_\sigma =& \{\sigma \in \mathbb{R}^+\}, \quad \textit{and}\\ \Theta_{\boldsymbol{\lambda}} =& \{\boldsymbol{\lambda} \in \mathbb{R}^d\}.\end{aligned}$$

The true parameter $\boldsymbol{\theta}_0$ *is an interior point of some compact and convex set* $\Theta_0 \subset \Theta$.

The following high level assumptions are used in showing the asymptotic results for the ML estimation, and can be found in Meitz and Saikkonen (2013).

Assumption 3.

A1. (i) *For all* $x \in \mathbb{R}$ *and* $\boldsymbol{\lambda} \in \Theta_{\boldsymbol{\lambda}}$, $f_\eta(x;\boldsymbol{\lambda})$ *is twice continuously differentiable w.r.t.* $(x, \boldsymbol{\lambda})$.

(ii) *For all* $\boldsymbol{\lambda} \in \Theta_{\boldsymbol{\lambda}}$, $\int x f_\eta(x;\boldsymbol{\lambda})dx = 0$ *and* $\int x^2 f_\eta(x;\boldsymbol{\lambda})dx = 1$.

(iii) *The matrix* $E[e_{\boldsymbol{\lambda},t}(\theta_0)e_{\boldsymbol{\lambda},t}(\theta_0)']$ *is positive definite.*

(iv) *For all* $x \in \mathbb{R}$ *and all* $\boldsymbol{\lambda}_i$, $i = 1, \ldots, d$*, functions*

$$x^4 \frac{f^2_{\eta,x}(x;\boldsymbol{\lambda}_0)}{f^2_\eta(x;\boldsymbol{\lambda}_0)} \quad \textit{and} \quad \frac{f^2_{\eta,\boldsymbol{\lambda}_i}(x;\boldsymbol{\lambda}_0)}{f^2_\eta(x;\boldsymbol{\lambda}_0)}$$

are dominated by $d_1(1+|x|^{d_2})$ *with some* $d_1, d_2 \geq 0$ *s.t.* $\int |x|^{d_2} f_\eta(x;\boldsymbol{\lambda}_0)dx < \infty$.

(v) *For all* $x \in \mathbb{R}$ *and* $\boldsymbol{\lambda} \in \Theta_{\boldsymbol{\lambda}}$*, the function* $|x^2 f_{\eta,\boldsymbol{\lambda}}(x;\boldsymbol{\lambda})|$ *is dominated by function* $\bar{f}(x)$ *s.t.* $\int \bar{f}(x)dx < \infty$.

A2. (i) *For all* $x \in \mathbb{R}$ *and* $\boldsymbol{\lambda} \in \Theta_{\boldsymbol{\lambda}}$*, the function* $|f_{\eta,\boldsymbol{\lambda}\boldsymbol{\lambda}}(x;\boldsymbol{\lambda})|$ *is dominated by some* $\bar{f}(x)$ *s.t.* $\int \bar{f}(x)dx < \infty$.

(ii) $\int f_{\eta,xx}(x;\boldsymbol{\lambda}_0)dx = 0$.

(iii) $\int x^2 f_{\eta,xx}(x;\boldsymbol{\lambda}_0)dx = 2$.

A3. (i) *For all* $x \in \mathbb{R}$ *and* $\boldsymbol{\lambda} \in \Theta_{\boldsymbol{\lambda}}$*, for all* $\boldsymbol{\lambda}_i$, $i = 1, \ldots, d$*, the functions*

$$x^4 \frac{f^4_{\eta,x}(x;\boldsymbol{\lambda})}{f^4_\eta(x;\boldsymbol{\lambda})}, \quad \frac{f^4_{\eta,\boldsymbol{\lambda}_i}(x;\boldsymbol{\lambda})}{f^4_\eta(x;\boldsymbol{\lambda})}, \quad x^4 \frac{f^2_{\eta,xx}(x;\boldsymbol{\lambda})}{f^4_\eta(x;\boldsymbol{\lambda})},$$

$$\frac{f^2_{\eta,\lambda_i x}(x;\boldsymbol{\lambda})}{f^4_\eta(x;\boldsymbol{\lambda})}, \quad \text{and} \quad \left|\frac{f_{\eta,\lambda\lambda}(x;\boldsymbol{\lambda})}{f_\eta(x;\boldsymbol{\lambda})}\right|$$

are dominated by $d_1(1+|x|^{d_2})$ *for some* $d_1, d_2 \geq 0$, *and* $\int |x|^{d_2} f_\eta(x;\boldsymbol{\lambda}_0)dx < \infty$.

A4. (i) *For all* $x \in \mathbb{R}$, $\Delta x \in \mathbb{R}$, *and* $\boldsymbol{\lambda} \in \Theta_\lambda$, *for some* $C < \infty$ *and* $d_1, d_2 \geq 0$,

$$|v(x+\Delta x;\boldsymbol{\lambda}) - v(x;\boldsymbol{\lambda})| \leq C((1+|x|^{d_1})|\Delta x| + |\Delta x|^{d_2})$$

for the following functions $v(x;\boldsymbol{\lambda})$,

$$\begin{aligned} &(i)\ v(x;\boldsymbol{\lambda}) = \frac{f_{\eta,x}(x;\boldsymbol{\lambda})}{f_\eta(x;\boldsymbol{\lambda})}, \quad (ii)\ v(x;\boldsymbol{\lambda}) = \frac{f_{\eta,\lambda}(x;\boldsymbol{\lambda})}{f_\eta(x;\boldsymbol{\lambda})}, \\ &(iii)\ v(x;\boldsymbol{\lambda}) = \frac{f_{\eta,\lambda\lambda}(x;\boldsymbol{\lambda})}{f_\eta(x;\boldsymbol{\lambda})}, \quad (iv)\ v(x;\boldsymbol{\lambda}) = \frac{f_{\eta,\lambda x}(x;\boldsymbol{\lambda})}{f_\eta(x;\boldsymbol{\lambda})}, \quad \text{and} \\ &(v)\ v(x;\boldsymbol{\lambda}) = \frac{f_{\eta,\lambda\lambda}(x;\boldsymbol{\lambda})}{f_\eta(x;\boldsymbol{\lambda})}. \end{aligned}$$

Assumptions 1, 2, and 3 are enough to state the following properties of (y_t, ε_t).

Lemma A1. *Under Assumption 1 (a) and 2, the process* (y_t, ε_t) *defined in* (2.1) *is stationary and ergodic. Moreover, the process* y_t *is* $\mathcal{F}_{t+Q}$*-measurable with* $E[y_t^4] < \infty$, *and* ε_t *is* $\mathcal{F}_t$*-measurable with* $E[\varepsilon_t^4] < \infty$. *If Assumption 1* (b) *also holds, then* $E[y_t^8] < \infty$ *and* $E[\varepsilon_t^8] < \infty$.

The proof will be omitted here, but essentially it can be found in Meitz and Saikkonen (2013) Appendix A, where series presentations of rational functions like $a(z,\boldsymbol{\theta})^{-1}$, $b(z^{-1},\boldsymbol{\theta})^{-1}$, $a(z,\boldsymbol{\theta})b(z^{-1},\boldsymbol{\theta})^{-1}$ and $a(z,\boldsymbol{\theta})^{-1}b(z^{-1},\boldsymbol{\theta})$ are discussed in depth. For future reference, we list the definitions of these sums here:

$$a(z,\boldsymbol{\theta})^{-1} = \sum_{j=0}^{\infty} \psi_j^{(a)} z^j, \quad b(z^{-1},\boldsymbol{\theta})^{-1} = \sum_{j=0}^{\infty} \psi_j^{(b)} z^{-j}$$

$$a(z,\boldsymbol{\theta})^{-1}b(z,\boldsymbol{\theta}) = \sum_{j=-P}^{\infty} \psi_j z^j \quad \text{and} \quad a(z,\boldsymbol{\theta})b(z^{-1},\boldsymbol{\theta})^{-1} = \sum_{j=-Q}^{\infty} \pi_j z^{-j}.$$

These series are well defined for all z in some area containing the unit circle, and the coefficients of the expansions are always geometrically decaying for all $\boldsymbol{\theta} \in \Theta$, given in Assumption 2.

2.B Derivatives of $u_t(\boldsymbol{\theta})$ and $\tilde{u}_t(\boldsymbol{\theta})$

The sequence $u_t(\boldsymbol{\theta})$ was defined in (2.2) and the discussion therein, and its feasible counterpart in (2.3). For what follows, we need a notion of the derivatives of these quantities. The derivative of $u_t(\boldsymbol{\theta})$ w.r.t. the p^{th} AR parameter and the q^{th} MA parameter are denoted by $u_{a_p,t}(\boldsymbol{\theta}) = \frac{\partial}{\partial b_q} u_t(\boldsymbol{\theta})$ and $u_{b_q,t}(\boldsymbol{\theta}) = \frac{\partial}{\partial b_q} u_t(\boldsymbol{\theta})$, respectively, for $p = 1, \ldots, P$ and $q = 1, \ldots, Q$. These functions are given by

$$u_{a_p,t}(\boldsymbol{\theta}) = -\frac{u_{t-p}(\boldsymbol{\theta})}{a(B)} = -\sum_{j=0}^{\infty} \psi_j^{(a)} u_{t-p-j}(\boldsymbol{\theta}) \quad \text{and}$$
$$u_{b_q,t}(\boldsymbol{\theta}) = \frac{u_{t+q}(\boldsymbol{\theta})}{b(B^{-1})} = \sum_{j=0}^{\infty} \psi_j^{(b)} u_{t+q+j}(\boldsymbol{\theta}).$$

Using a representation $\tilde{u}_t(\boldsymbol{\theta}) = \sum_{j=0}^{T-t} \psi_j^{(b)} a(B) y_{t+j}$ (Andrews et al., 2006), the derivatives of the feasible quantities $\tilde{u}_t(\boldsymbol{\theta})$ are given by

$$\tilde{u}_{a_p,t}(\boldsymbol{\theta}) = -\sum_{j=0}^{T-t} \psi_j^{(b)} y_{t-p-j} \quad \text{and} \quad \tilde{u}_{b_q,t}(\boldsymbol{\theta}) = \sum_{j=0}^{T-t} \psi^{(b)j} \tilde{u}_{t+q+j}(\boldsymbol{\theta}).$$

For convenience, the P and Q vectors of the derivatives of $u_t(\boldsymbol{\theta})$ are denoted by

$$\frac{\partial}{\partial \boldsymbol{\theta}_a} u_t(\boldsymbol{\theta}) = \boldsymbol{u}_{a,t}(\boldsymbol{\theta}) = \begin{pmatrix} u_{a_1,t}(\boldsymbol{\theta}) \\ \vdots \\ \boldsymbol{u}_{a_P,t}(\boldsymbol{\theta}) \end{pmatrix} \quad \text{and} \quad \frac{\partial}{\partial \boldsymbol{\theta}_b} u_t(\boldsymbol{\theta}) = \begin{pmatrix} u_{b_1,t}(\boldsymbol{\theta}) \\ \vdots \\ u_{b_Q,t}(\boldsymbol{\theta}) \end{pmatrix},$$

and respectively for $\tilde{u}_{a,t}(\boldsymbol{\theta})$ and $\tilde{u}_{b,t}(\boldsymbol{\theta})$ in an obvious manner.

2.C Intermediate results for the model

In this section we provide some details for the proofs of the results presented in the main text. The approximation of the log-likelihood function was presented in (2.4). Its feasible counterpart is

$$\tilde{L}_T(\boldsymbol{\theta}) = T^{-1} \sum_{t=1}^{T} \tilde{l}_t(\boldsymbol{\theta}) \quad \text{with} \quad \tilde{l}_t = \log f_\eta \left(\sigma^{-1}\tilde{u}_t(\boldsymbol{\theta}); \boldsymbol{\lambda}\right) - \log \sigma.$$

The $(P + Q + 1 + d)$ dimensional score vector of a single observation at $\boldsymbol{\theta}$ is denoted by $l_{\boldsymbol{\theta},t}(\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} l_t(\boldsymbol{\theta})$, and it is

$$l_{\boldsymbol{\theta},t}(\boldsymbol{\theta}) = \begin{pmatrix} \sigma^{-1} e_{x,t}(\boldsymbol{\theta}) \boldsymbol{u}_{a,t}(\boldsymbol{\theta}) \\ \sigma^{-1} e_{x,t}(\boldsymbol{\theta}) \boldsymbol{u}_{b,t}(\boldsymbol{\theta}) \\ -\sigma^{-1}\left(\sigma^{-1} e_{x,t}(\boldsymbol{\theta}) u_t(\boldsymbol{\theta}) + 1\right) \\ e_{\boldsymbol{\lambda},t}(\boldsymbol{\theta}) \end{pmatrix}.$$

The score vector of the model is given by $L_{\boldsymbol{\theta},T}(\boldsymbol{\theta}) = T^{-1} \sum_{t=1}^{T} l_{\boldsymbol{\theta},t}(\boldsymbol{\theta})$.

The Hessian of the noninvertible ARMA model is substantially more involved than that of the invertible ARMA model (although it is simplified substantially from the Hessian presented in Meitz and Saikkonen (2013) as we neglect the ARCH error term). It is not shown here, but one can confirm that

$$-\mathrm{E}[l_{\boldsymbol{\theta}\boldsymbol{\theta}',t}(\boldsymbol{\theta}_0)] = \lim_{T\to\infty} \mathrm{Cov}\left(T^{-1/2} \sum_{t=1}^{T} l_{\boldsymbol{\theta},t}(\boldsymbol{\theta}_0)\right),$$

where the limit is positive definite, continuous, and finite in Θ_0, given in Assumption 2. This matrix is (see Proposition 1)

$$\boldsymbol{\mathcal{I}} = \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{B}_{21}' & \boldsymbol{0}_{P\times 1} & \boldsymbol{0}_{P\times d} \\ \boldsymbol{B}_{21} & \boldsymbol{A}_{22} & \boldsymbol{0}_{Q\times 1} & \boldsymbol{0}_{Q\times d} \\ \boldsymbol{0}_{1\times P} & \boldsymbol{0}_{1\times Q} & \boldsymbol{A}_{33} & \boldsymbol{A}_{43}' \\ \boldsymbol{0}_{d\times P} & \boldsymbol{0}_{d\times Q} & \boldsymbol{A}_{43} & \boldsymbol{A}_{44} \end{pmatrix}. \tag{2.7}$$

Straightforward but rather long calculations give

$$\boldsymbol{A}_{11} = -\,\sigma_0^{-2}\mathrm{E}[e_{x,t}^2]\mathrm{E}[\boldsymbol{u}_{\boldsymbol{\theta}_a,t}(\boldsymbol{\theta_0})\boldsymbol{u}_{\boldsymbol{\theta}'_a,t}(\boldsymbol{\theta}_0)], \quad \boldsymbol{A}_{22} = \sigma_0^{-2}\mathrm{E}[e_{x,t}^2]\mathrm{E}[\boldsymbol{u}_{\boldsymbol{\theta}_b,t}(\boldsymbol{\theta_0})\boldsymbol{u}_{\boldsymbol{\theta}'_b,t}(\boldsymbol{\theta_0})],$$

$$\boldsymbol{A}_{33} = \sigma_0^{-2}(\mathrm{E}[e_{x,t}^2\eta_t^2]-1)^2, \quad \boldsymbol{A}_{43} = -\sigma_0^{-1}\mathrm{E}[e_{x,t}e_{\boldsymbol{\lambda},t}\eta_t], \quad \boldsymbol{A}_{44} = -\mathrm{E}[e_{\boldsymbol{\lambda},t}e_{\boldsymbol{\lambda}',t}], \quad \text{and}$$

$$\boldsymbol{B}_{21} = -\,\sigma_0^{-2}\mathrm{E}[\boldsymbol{u}_{\boldsymbol{\theta}_a,t}(\boldsymbol{\theta_0})\boldsymbol{u}_{\boldsymbol{\theta}'_b,t}(\boldsymbol{\theta}_0)].$$

The block $\boldsymbol{B}_{21}$ is due to the serial correlation of the score vector whereas the the rest of the blocks capture the contemporaneous correlation. The expressions above have feasible counterparts obtained by setting

$$\tilde{e}_{x,t} = \frac{f_{\eta,x}(\sigma_0^{-1}\tilde{u}_t(\boldsymbol{\theta}_0);\boldsymbol{\lambda}_0)}{f_\eta(\sigma_0^{-1}\tilde{u}_t(\boldsymbol{\theta}_0);\boldsymbol{\lambda}_0)}, \quad \text{and} \quad \tilde{e}_{\boldsymbol{\lambda},t}(\boldsymbol{\theta}_0) = \frac{f_{\eta,\boldsymbol{\lambda}}(\sigma_0^{-1}\tilde{u}_t(\boldsymbol{\theta}_0);\boldsymbol{\lambda}_0)}{f_\eta(\sigma_0^{-1}\tilde{u}_t(\boldsymbol{\theta}_0);\boldsymbol{\lambda}_0)}.$$

The set of results in Lemma C2 is used by Meitz and Saikkonen (2013) to derive the results we presented in Proposition 1, but it is also needed in the proof of Lemma C3 and Lemma C4, and Lemmas D5-D7.

Lemma C2. *Under Assumptions 1 (a), 2 and 3,*

$$(i) \left\| \sup_{\boldsymbol{\theta}\in\Theta_0} |u_t(\boldsymbol{\theta})| \right\|_4 < \infty, \quad (ii) \left\| \sup_{\boldsymbol{\theta}\in\Theta_0} \left| \frac{u_t(\boldsymbol{\theta})}{a(B)} \right| \right\|_4 < \infty,$$

$$(iii) \left\| \sup_{\boldsymbol{\theta}\in\Theta_0} \left| \frac{u_t(\boldsymbol{\theta})}{b(B^{-1})} \right| \right\|_4 < \infty, \quad (iv) \left\| \sup_{\boldsymbol{\theta}\in\Theta_0} |e_{\boldsymbol{\lambda},t}(\boldsymbol{\theta})| \right\|_2 < \infty,$$

$$(v) \left\| \sup_{\boldsymbol{\theta}\in\Theta_0} |e_{x,t}(\boldsymbol{\theta})| \right\|_2 < \infty, \quad (vi)\ E[e_{x,t}] = 0,$$

$$(vii)\ E[e_{x,t}\varepsilon_t] = -\sigma_0, \quad (viii)\ E[e_{x,t}\varepsilon_t^2] = 0, \quad (ix)\ E[e_{x,t}\varepsilon_t^3] = -3\sigma_0^3,$$

$$(x)\ E[e_{\boldsymbol{\lambda},t}] = 0, \quad (xi)\ E[e_{\boldsymbol{\lambda},t}\varepsilon_t^2] = 0.$$

Proof. Proofs for the results in this section are given in a supplementary appendix. □

The next lemma provides some insight between the feasible and unfeasible quantities.

Lemma C3. *Under Assumptions 1 (a), 2, and 3,*

$$(i)\ \left\| \sup_{\boldsymbol{\theta}\in\Theta_0} |u_t(\boldsymbol{\theta}) - \tilde{u}_t(\boldsymbol{\theta})| \right\|_4 \leq CK_t, \quad (ii)\ \left\| \sup_{\boldsymbol{\theta}\in\Theta_0} |u_{a_p,t}(\boldsymbol{\theta}) - \tilde{u}_{a_p,t}(\boldsymbol{\theta})| \right\|_4 \leq CK_t,$$

$$(iii)\ \left\| \sup_{\boldsymbol{\theta}\in\Theta_0} |u_{b_p,t}(\boldsymbol{\theta}) - \tilde{u}_{b_p,t}(\boldsymbol{\theta})| \right\|_4 \leq CK_t,\ (iv)\ \left\| \sup_{\boldsymbol{\theta}\in\Theta_0} |e_{x,t}(\boldsymbol{\theta}) - \tilde{e}_{x,t}(\boldsymbol{\theta})| \right\|_{r_1} \leq CK_t,$$

$$(v)\ \left\| \sup_{\boldsymbol{\theta}\in\Theta_0} |u_t(\boldsymbol{\theta})^2 - \tilde{u}_t(\boldsymbol{\theta})^2| \right\|_2 \leq CK_t,$$

where $C < \infty$ is a constant and may vary from part to part, K_t is a constant that may depend on t and decays to zero at geometric rate as $T \to \infty$, and r_1 is some small positive number.

The following results are used directly in the proofs of Lemmas 1 and 3.

Lemma C4. *Under Assumptions 1 (a), 2, and 3, as $T \to \infty$,*

$$(i)\ T^{1/2} \sup_{\boldsymbol{\theta}\in\Theta_0} \left| L_{\boldsymbol{\theta},T}(\boldsymbol{\theta}) - \tilde{L}_{\boldsymbol{\theta},T}(\boldsymbol{\theta}) \right| \stackrel{a.s.}{\to} \mathbf{0} \quad \text{and } (ii)\ \sup_{\boldsymbol{\theta}\in\Theta_0} \left| L_{\boldsymbol{\theta}\boldsymbol{\theta}',T}(\boldsymbol{\theta}) - \tilde{L}_{\boldsymbol{\theta}\boldsymbol{\theta}',T}(\boldsymbol{\theta}) \right| \stackrel{a.s.}{\to} \mathbf{0}.$$

2.D Intermediate results for the test statistics

The next three Lemmas are used in the proofs of Lemmas 1-4. Lemma D5 gives some essential uniform convergence results for the unfeasible autocorrelation functions. Lemma D6 justifies the asymptotic normal distribution of the vectors $(L_{\boldsymbol{\theta},T}(\boldsymbol{\theta}_0), \gamma_{ac})$ and $(L_{\boldsymbol{\theta},T}(\boldsymbol{\theta}_0), \gamma_{hs})$ that we encounter in Lemmas 2 and 4. Lemma D7 is used in the proofs of Lemmas 1-4 to ensure that the asymptotic properties hold for the feasible quantities as well as the unfeasible ones.

For what follows, let $\boldsymbol{g}_{ac,t}(\boldsymbol{\theta}) = (u_t(\boldsymbol{\theta})(u_{t-1}(\boldsymbol{\theta}), \ldots, u_{t-m}(\boldsymbol{\theta})))$ and $\boldsymbol{g}_{hs,t}(\boldsymbol{\theta}) = (u_t(\boldsymbol{\theta})^2 - \sigma^2)(u_{t-1}(\boldsymbol{\theta})^2 - \sigma^2, \ldots, u_{t-m}(\boldsymbol{\theta})^2 - \sigma^2)$.

Lemma D5. *Under Assumptions 1 (a), 2, and 3, for $j = 1\ldots, m$, as $T \to \infty$,*

$$(i)\ \sup_{\boldsymbol{\theta}\in\Theta_0}\left|\frac{\partial}{\partial\boldsymbol{\theta}'}\gamma_{ac}(\boldsymbol{\theta}) - \boldsymbol{G}_{ac}(\boldsymbol{\theta})\right| \overset{a.s.}{\to} \mathbf{0}, \quad and \quad (ii)\ \sup_{\boldsymbol{\theta}\in\Theta_0}\left|\frac{\partial}{\partial\boldsymbol{\theta}'}\gamma_{hs}(\boldsymbol{\theta}) - \boldsymbol{G}_{hs}(\boldsymbol{\theta})\right| \overset{a.s.}{\to} \mathbf{0},$$

where constant matrices $\boldsymbol{G}_{ac}(\boldsymbol{\theta}) = E[\frac{\partial}{\partial\boldsymbol{\theta}'}g_{ac,t}(\boldsymbol{\theta})]$ and $\boldsymbol{G}_{hs}(\boldsymbol{\theta}) = E[\frac{\partial}{\partial\boldsymbol{\theta}'}g_{hs,t}(\boldsymbol{\theta})]$ are finite and continuous. Moreover, $\boldsymbol{G}_{ac}(\boldsymbol{\theta_0}) = \boldsymbol{G}_{ac} = -\boldsymbol{\Sigma}_{l,\gamma_{ac}}$ given in Lemma 2 and (2.5), *and $\boldsymbol{G}_{hs}(\boldsymbol{\theta_0}) = \boldsymbol{G}_{hs} = \mathbf{0}_{m\times(P+Q+1+d)}$.*

Proof. Results follow from Theorem 2.7. in Straumann and Mikosch (2006), if we show that the summands $g_{ac,t}(\boldsymbol{\theta})$ and $g_{hs,t}(\boldsymbol{\theta})$ are bounded in a sup-norm sense: $\|\sup_{\boldsymbol{\theta}\in\Theta_0}|g_{ac,t}(\boldsymbol{\theta})|\| < \infty$ and $\|\sup_{\boldsymbol{\theta}\in\Theta_0}|g_{hs,t}(\boldsymbol{\theta})|\| < \infty$. Details can be found in the supplementary appendix. Exact forms of the matrices $\boldsymbol{G}_{ac}$ and $\boldsymbol{G}_{hs}$ are straightforward to verify and the details can be found in the supplementary appendix. □

Lemma D6. *Let $\boldsymbol{\xi}_{ac,t}(\boldsymbol{\theta}) = (l_t(\boldsymbol{\theta}), g_{ac,t}(\boldsymbol{\theta}))$ and $\boldsymbol{\xi}_{hs,t}(\boldsymbol{\theta}) = (l_t(\boldsymbol{\theta}), g_{hs,t}(\boldsymbol{\theta}))$, then*

A under Assumptions 1 (a), 2, and 3,

- *(i) vectors $\boldsymbol{\xi}_{ac,t}(\boldsymbol{\theta}_0)$ and $\boldsymbol{\xi}_{hs,t}(\boldsymbol{\theta}_0)$ form a stationary and ergodic process with $E[\boldsymbol{\xi}_{ac,t}(\boldsymbol{\theta}_0)] = E[\boldsymbol{\xi}_{hs,t}(\boldsymbol{\theta}_0)] = \mathbf{0}$,*
- *(ii) $\boldsymbol{\xi}_{ac,t}(\boldsymbol{\theta}_0)$ has a finite covariance matrix $E[\boldsymbol{\xi}_{ac,t}(\boldsymbol{\theta}_0)\boldsymbol{\xi}_{ac,t}(\boldsymbol{\theta}_0)'] < \infty$,*
- *(iii) for all conformable nonrandom vectors $\boldsymbol{a} \neq \mathbf{0}$, sequences $\boldsymbol{a}'\boldsymbol{\xi}_{ac,t}(\boldsymbol{\theta}_0)$ and $\boldsymbol{a}'\boldsymbol{\xi}_{hs,t}(\boldsymbol{\theta}_0)$ are L_2-mixingales of size -1 w.r.t. the filtration $\{\mathcal{F}_s\}_{s\leq t}$, and*
- *(iv) there is a finite and positive definite limiting covariance matrix $\boldsymbol{\Sigma}_{\xi_{ac}}$ s.t.*

$$\lim_{T\to\infty} Cov\left(T^{-1/2}\sum_{t=1}^{T}\boldsymbol{\xi}_{ac,t}(\boldsymbol{\theta}_0)\right) \overset{a.s.}{\to} \boldsymbol{\Sigma}_{\xi_{ac}}.$$

B under Assumptions 1 (b), 2, and 3,

- *(i) $\boldsymbol{\xi}_{hs,t}(\boldsymbol{\theta}_0)$ has a finite covariance matrix $E[\boldsymbol{\xi}_{hs,t}(\boldsymbol{\theta}_0)\boldsymbol{\xi}_{hs,t}(\boldsymbol{\theta}_0)'] < \infty$, and*
- *(ii) there is a finite and positive definite limiting covariance matrix $\boldsymbol{\Sigma}_{\xi_{hs}}$ s.t.*

$$\lim_{T\to\infty} Cov\left(T^{-1/2}\sum_{t=1}^{T}\boldsymbol{\xi}_{hs,t}(\boldsymbol{\theta}_0)\right) \overset{a.s.}{\to} \boldsymbol{\Sigma}_{\xi_{hs}}.$$

Proof. Part A (i) holds because vectors $\boldsymbol{\xi}_{ac,t}(\boldsymbol{\theta}_0)$ and $\boldsymbol{\xi}_{hs,t}(\boldsymbol{\theta}_0)$ can be expressed in terms of convergent power series expansions of the ergodic and stationary process $u_t(\boldsymbol{\theta}_0) = \varepsilon_t$. A (ii) is a consequence of A (iv). A (iii) is trivial, since, by Lemma 3 in Meitz and Saikkonen (2013), for a conformable size vector $\boldsymbol{a}_1 \neq \mathbf{0}$, $\boldsymbol{a}_1' l_t(\boldsymbol{\theta}_0)$ is an L_2-mixingale of size -1, and Vectors $\boldsymbol{g}_{ac,t}(\boldsymbol{\theta}_0)$ and $\boldsymbol{g}_{hs,t}(\boldsymbol{\theta}_0)$ are $\mathcal{F}_t$-measurable mean zero processes. Part B (i) is a consequence of B (ii). Rather long derivations for the proofs of parts A (iv) and B (ii) can be found in the supplementary appendix. The exact forms of the matrices $\boldsymbol{\Sigma}_{\xi_{ac}}$ and $\boldsymbol{\Sigma}_{\xi_{hs}}$ are presented in Lemmas 1 and 3. For the details, see the proofs of A (iv) and B (ii) in the supplementary appendix. □

Proposition 2. *(i) Under Assumptions 1 (a), 2, and 3, $T^{1/2}(L_{\boldsymbol{\theta},T}(\boldsymbol{\theta}_0), \gamma_{ac}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}_{\xi_{ac}})$, and (ii) under Assumptions 1 (b), 2, and 3, $T^{1/2}(L_{\boldsymbol{\theta},T}(\boldsymbol{\theta}_0), \gamma_{hs}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}_{\xi_{hs}})$.*

Proof. A direct application of Lemma D6 and Lemma A.4. in Meitz and Saikkonen (2013) gives $T^{1/2}\boldsymbol{a}'(L_{\boldsymbol{\theta},T}(\boldsymbol{\theta}_0), \gamma_{ac}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{a}'\boldsymbol{\Sigma}_{\xi_{ac}}\boldsymbol{a})$ and $T^{1/2}\boldsymbol{a}'(L_{\boldsymbol{\theta},T}(\boldsymbol{\theta}_0), \gamma_{hs}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{a}'\boldsymbol{\Sigma}_{\xi_{hs}}\boldsymbol{a})$, for all conformable size vectors $\boldsymbol{a} \neq 0$. Proposition 2 follows by Cramér-Wold device. □

Lemma D7. *Under Assumptions 1 (a), 2, and 3, as $T \to \infty$*

$$(i) \sup_{\boldsymbol{\theta}\in\Theta_0} |T^{1/2}\gamma_{ac}(\boldsymbol{\theta}) - T^{1/2}\tilde{\gamma}_{ac}(\boldsymbol{\theta})| \xrightarrow{a.s.} \mathbf{0},$$

$$(ii) \sup_{\boldsymbol{\theta}\in\Theta_0} |T^{1/2}\gamma_{hs}(\boldsymbol{\theta}) - T^{1/2}\tilde{\gamma}_{hs}(\boldsymbol{\theta})| \xrightarrow{a.s.} \mathbf{0}$$

$$(iii) \sup_{\boldsymbol{\theta}\in\Theta_0} \left| \frac{\partial}{\partial\boldsymbol{\theta}'}\gamma_{ac}(\boldsymbol{\theta}) - \frac{\partial}{\partial\boldsymbol{\theta}'}\tilde{\gamma}_{ac}(\boldsymbol{\theta}) \right| \xrightarrow{a.s.} \mathbf{0}, \quad \textit{and}$$

$$(iv) \sup_{\boldsymbol{\theta}\in\Theta_0} \left| \frac{\partial}{\partial\boldsymbol{\theta}'}\gamma_{hs}(\boldsymbol{\theta}) - \frac{\partial}{\partial\boldsymbol{\theta}'}\tilde{\gamma}_{hs}(\boldsymbol{\theta}) \right| \xrightarrow{a.s.} \mathbf{0}.$$

Proof. Results follow by Theorem 2.7. in Straumann and Mikosch (2006), and Lemma C3. Details can be found in the supplementary appendix. □

Supplementary appendix for the Residual-based diagnostic tests for noninvertible ARMA models

Proof of Lemma C2.

(i) Recall the representation (2.2), $u_t(\boldsymbol{\theta}) = \sum_{j=-P}^{\infty} \pi_j y_{t+j}$. By Lemma A1 (i), $\mathrm{E}[y_t^4] < \infty$. The result follows by Lemmas A.1. and A.2. in Meitz and Saikkonen (2013) (MS13 hereafter). (ii) Because $a(B)^{-1}u_t(\boldsymbol{\theta}) = \sum_{j=0}^{\infty} \psi_j^{(a)} u_{t-j}(\boldsymbol{\theta})$, result follows from part (i) and Lemmas A.1. and A.2. in MS13. (iii) This is analogous to part (ii). Parts (iv) and (v) follows from the continuous differentiability of $f_\eta(x;\boldsymbol{\lambda})$ w.r.t. $(x,\boldsymbol{\lambda})$. Parts $(vi)-(xi)$ can be found in Lemma C.1. in MS13.

Proof of Lemma C3.

Proof of Lemma C3 is very close to the Lemma E.1. in MS13. (i) Using the presentation $u_t(\boldsymbol{\theta}) = a(B)b(B^{-1})^{-1}y_t = \sum_{j=0}^{\infty} \psi_j^{(b)} a(B)y_{t+j}$ together with $\tilde{u}_t(\boldsymbol{\theta}) = \sum_{j=0}^{T-t} \psi_j^{(b)} a(B)y_{t+j}$ gives

$$u_t(\boldsymbol{\theta}) - \tilde{u}_t(\boldsymbol{\theta}) = \sum_{j=T-t+1}^{\infty} \psi_j^{(b)} a(B)y_{t+j}.$$

Triangle inequality and Hölder's inequality implies

$$\left\| \sup_{\boldsymbol{\theta}\in\boldsymbol{\theta}_0} |u_t(\boldsymbol{\theta}) - \tilde{u}_t(\boldsymbol{\theta})| \right\|_4 \leq \sum_{j=T-t+1}^{\infty} |\psi_j^{(b)}| \left\| \sup_{\boldsymbol{\theta}\in\Theta_0} |a(B)y_{t+j}| \right\|_4 .$$

As the Laurent series coefficients $\psi_j^{(b)}$ are geometrically decaying, there are constants $C_1 < \infty$ and $|\rho| < 1$ s.t. $\psi_j^{(b)} \leq C_1\rho^j$. Finite fourth moments of process y_t, together with compactness of $\boldsymbol{\Theta 0}, \boldsymbol{a}$, ensures that $\left\| \sup_{\boldsymbol{\theta}\in\Theta_0} |a(B)y_{t+j}(\boldsymbol{\theta})| \right\|_4 \leq C_2$, for some constant $C_2 < \infty$. Using these, we get

$$\left\| \sup_{\boldsymbol{\theta}\in\Theta_0} |u_t(\boldsymbol{\theta}) - \tilde{u}_t(\boldsymbol{\theta})| \right\|_4 \leq \sum_{j=T-t+1}^{\infty} \rho^j C_1 C_2 = \frac{\rho^{T-t+1}}{1-\rho} C_1 C_2,$$

and the result follows.

(*ii*) Recall expressions $u_{a_p,t}(\boldsymbol{\theta}) = -\sum_{j=0}^{\infty}\psi_j^{(a)}u_{t-p-j}(\boldsymbol{\theta})$ and $\tilde{u}_{a_p,t}(\boldsymbol{\theta}) = -\sum_{j=0}^{T-t}\psi_j^{(b)}y_{t-p+j}$ in Appendix 2.B, and note that

$$\left\|\sup_{\boldsymbol{\theta}\in\Theta_0}\left|u_{a_p,t}(\boldsymbol{\theta})-\tilde{u}_{a_p,t}(\boldsymbol{\theta})\right|\right\|_4 \leq \sum_{j=T-t+1}^{\infty}\sup_{\boldsymbol{\theta}\in\Theta_0}|\psi_j^{(b)}|\left\|y_{t-p+j}\right\|_4.$$

Because $|\sup_{\boldsymbol{\theta}\in\Theta_0}\psi_j^{(b)}| \leq C_1\rho^j$, and process y_t has finite fourth moment, $\mathrm{E}[y_t^4] \leq C_2 < \infty$,

$$\left\|\sup_{\boldsymbol{\theta}\in\Theta_0}|u_{a_p,t}(\boldsymbol{\theta})-\tilde{u}_{a_p,t}(\boldsymbol{\theta})|\right\| \leq \sum_{j=T-t+1}^{\infty}\rho^j C_1C_2 = \frac{\rho^{T-t+1}}{1-\rho}C_1C_2.$$

(*iii*) Recall expressions $u_{b_q,t}(\boldsymbol{\theta}) = \sum_{j=0}^{\infty}\psi_j^{(b)}u_{t+q+j}(\boldsymbol{\theta})$ and $\tilde{u}_{b_q,t}(\boldsymbol{\theta}) = \sum_{j=0}^{T-t}\psi_j^{(b)}\tilde{u}_{t+q+j}(\boldsymbol{\theta})$ in Appendix 2.B,

$$\begin{aligned}
|u_{b_q,t}(\boldsymbol{\theta})-\tilde{u}_{b_q,t}(\boldsymbol{\theta})| &= \left|\sum_{j=0}^{\infty}\psi_j^{(b)}u_{t+q+j} - \sum_{j=0}^{T-t}\psi_j^{(b)}\tilde{u}_{t+q+j}(\boldsymbol{\theta})\right| \\
&= \left|\sum_{j=0}^{\infty}\psi_j^{(b)}u_{t+q+j}(\boldsymbol{\theta}) - \sum_{j=0}^{T-t}\psi_j^{(b)}u_{t+q+j}(\boldsymbol{\theta})\right. \\
&\quad \left. + \sum_{j=0}^{T-t}\psi_j^{(b)}u_{t+q+j}(\boldsymbol{\theta}) - \sum_{j=0}^{T-t}\psi_j^{(b)}\tilde{u}_{t+q+j}(\boldsymbol{\theta})\right| \\
&\leq \sum_{j=T-t+1}^{\infty}|\psi_j^{(b)}u_{t+q+j}(\boldsymbol{\theta})| \\
&\quad + \sum_{j=0}^{T-t}|\psi_j^{(b)}||u_{t+q+j}(\boldsymbol{\theta})-\tilde{u}_{t+q+j}(\boldsymbol{\theta})| \\
&= \sum_{j=T-t+1}^{\infty}|\psi_j^{(b)}u_{t+q+j}(\boldsymbol{\theta})| \\
&\quad + \sum_{j=0}^{T-t-q}|\psi_j^{(b)}||u_{t+q+j}(\boldsymbol{\theta})-\tilde{u}_{t+q+j}(\boldsymbol{\theta})|
\end{aligned}$$

$$+ \sum_{T-t-q+1}^{T-t} |\psi_j^{(b)} u_{t+q+j}(\boldsymbol{\theta})|,$$

where the first equality follows by adding and subtracting terms, the first inequality follows by the triangle inequality, and the last equality follows, because for $t > T$, $\tilde{u}_t(\boldsymbol{\theta}) = 0$. Taking sup-norm on both sides gives

$$\begin{aligned}\left\| \sup_{\boldsymbol{\theta}\in\Theta_0} |u_{b_q,t}(\boldsymbol{\theta}) - \tilde{u}_{b_q,t}(\boldsymbol{\theta})| \right\|_4 \leq & \sum_{j=T-t+1}^{\infty} \sup_{\boldsymbol{\theta}\in\Theta} |\psi_j^{(b)}| \left\| \sup_{\boldsymbol{\theta}\in\Theta_0} |u_{t+q+j}(\boldsymbol{\theta})| \right\|_4 \\ & + \sum_{j=0}^{T-t-q} \sup_{\boldsymbol{\theta}\in\Theta_0} |\psi_j^{(b)}| \left\| \sup_{\boldsymbol{\theta}\in\Theta_0} |u_{t+q+j}(\boldsymbol{\theta}) - \tilde{u}_{t+q+j}(\boldsymbol{\theta})| \right\|_4 \\ & + \sum_{j=T-t-q+1}^{T-t} \sup_{\boldsymbol{\theta}\in\Theta_0} |\psi_j^{(b)}| \left\| \sup_{\boldsymbol{\theta}\in\Theta_0} |u_{t+q+j}(\boldsymbol{\theta})| \right\|_4 .\end{aligned}$$

Using Lemma C2, the first term is bounded by

$$\sum_{T-t+1}^{\infty} \rho^j C_1 C_2 = \frac{\rho^{T-t+1}}{1-\rho} C_1 C_2.$$

From part (i), the second term can be bounded by

$$\sum_{j=0}^{T-t-q} C_3 \rho^j \frac{\rho^{T-t-q-j+1}}{1-\rho} C_4 C_5 \leq (T-t+1) \frac{\rho^{T-t+1}}{1-\rho} C_6.$$

The last term is bounded by

$$\sum_{j=T-t-q+1}^{T-t} C_7 \rho^j C_8 = C_7 C_8 \frac{\rho^{T-t-q+1} - \rho^{T-t+1}}{1-\rho} = \frac{\rho^{T-t+1}}{1-\rho} C_9.$$

Combining these yields the result,

$$\left\| \sup_{\boldsymbol{\theta}\in\Theta_0} |u_{b_q,t}(\boldsymbol{\theta}) - \tilde{u}_{b_q,t}(\boldsymbol{\theta})|| \right\|_4 \leq (T-t+1) \frac{\rho^{T-t+1}}{1-\rho} C,$$

where $C < \infty$ is not dependent on T or t.

(iv) Assumptions 3 A4 (i) directly implies

$$\begin{aligned}&|e_{x,t}(\boldsymbol{\theta}) - \tilde{e}_{x,t}(\boldsymbol{\theta})|\\ \leq& C((1+|\sigma^{-1}u_t(\boldsymbol{\theta})|^{d_1})\sigma^{-1}|u_t(\boldsymbol{\theta}) - \tilde{u}_t(\boldsymbol{\theta})| + (\sigma^{-1}|u_t(\boldsymbol{\theta}) - \tilde{u}_t(\boldsymbol{\theta})|)^{d_2}).\end{aligned}$$

Taking sup-norms on both sides yields

$$\begin{aligned}\left\|\sup_{\boldsymbol{\theta}\in\Theta_0} e_{x,t}(\boldsymbol{\theta}) - \tilde{e}_{x,t}(\boldsymbol{\theta})|\right\| \leq& C_1\left[\left\|\sup_{\boldsymbol{\theta}\in\Theta_0}|(1+u_t(\boldsymbol{\theta})^{d_1})|u_t(\boldsymbol{\theta}) - \tilde{u}_t(\boldsymbol{\theta})||\right\|_r\right.\\ &\left.+\left\|\sup_{\boldsymbol{\theta}\in\Theta_0}|u_t(\boldsymbol{\theta}) - \tilde{u}_t(\boldsymbol{\theta})|^{d_2}\right\|_r\right]\\ \leq& C_1\left[\left\|\sup_{\boldsymbol{\theta}\in\Theta_0}|1+u_t(\boldsymbol{\theta})^{d_1}|\right\|_{2r}\left\|\sup_{\boldsymbol{\theta}\in\Theta_0}|u_t(\boldsymbol{\theta}) - \tilde{u}_t(\boldsymbol{\theta})|\right\|_{2r}\right.\\ &\left.+\left\|\sup_{\boldsymbol{\theta}\in\Theta_0}|u_t(\boldsymbol{\theta}) - \tilde{u}_t(\boldsymbol{\theta})|^{d_2}\right\|_r\right]\end{aligned}$$

where the first inequality follows from Lòeves inequality and the second is due to Cauchy-Schwartz inequality. Now, choose r s.t. $r \leq \min\{2/d_1, 2, 4/d_2\}$, and note that the first term is bounded by some constant C_2 by Lemma C2 (i), and the second and the third are bounded by part (i), so

$$\begin{aligned}\left\|\sup_{\boldsymbol{\theta}\in\Theta_0}|e_{x,t}(\boldsymbol{\theta}) - \tilde{e}_{x,t}(\boldsymbol{\theta})|\right\|_r \leq& C_1\left[C_2\frac{\rho^{T-t+1}}{1-\rho}C_3C_4 + \frac{\rho^{T-t+1}}{1-\rho}C_5C_6\right]\\ =& C_7\frac{\rho^{T-t+1}}{1-\rho}.\end{aligned}$$

(v) Applying inequality $|x^2 - z^2| \leq |x-z|^2 + 2|x||x-z|$, the sup-norm is bounded by

$$\begin{aligned}&\leq \left\|\sup_{\boldsymbol{\theta}\in\Theta_0}|u_t(\boldsymbol{\theta}) - \tilde{u}_t(\boldsymbol{\theta})|\right\|_4^2 + 2\left\|\sup_{\boldsymbol{\theta}\in\Theta_0}|u_t(\boldsymbol{\theta})|\right\|_4\left\|\sup_{\boldsymbol{\theta}\in\Theta_0}|u_t(\boldsymbol{\theta}) - \tilde{u}_t(\boldsymbol{\theta})|\right\|_4\\ &\left(\frac{\rho^{T-t+1}}{1-\rho}C_1C_2\right)^2 + C_3\frac{\rho^{T-t+1}}{1-\rho}C_1C_2.\end{aligned}$$

Proof of Lemma C4

Part (i) is a simplified versions of Lemma 7 (ii) in MS13. We demonstrate the proof of part (i) for the element $|L_{b_q,T}(\boldsymbol{\theta}) - \tilde{L}_{b_q,T}(\boldsymbol{\theta})|$. The same technique applies to all the elements in that vector.

The line of the proof is the following. First, use the triangle inequality to obtain

$$T^{1/2} \sup_{b_q \in \Theta_{0,b_q}} |L_{b_q,T}(\boldsymbol{\theta}) - \tilde{L}_{b_q,T}(\boldsymbol{\theta})| \leq T^{-1/2} \sum_{t=1}^{T} \sup_{b_q \in \Theta_{0,b_q}} |l_{b_q,t}(\boldsymbol{\theta}) - \tilde{l}_{b_q,t}(\boldsymbol{\theta})|. \quad (2.8)$$

The result follows by justifying that the sum on the majorant side has a finite limit as $T \to \infty$. In order to do so, note that

$$\sup_{\boldsymbol{\theta} \in \Theta_0} |l_{b_q,t}(\boldsymbol{\theta}) - \tilde{l}_{b_q,t}(\boldsymbol{\theta})| = \sup_{\boldsymbol{\theta} \in \Theta_0} |\sigma^{-1}(|e_{x,t}(\boldsymbol{\theta}) u_{b_q,t}(\boldsymbol{\theta}) - \tilde{e}_{x,t}(\boldsymbol{\theta}) \tilde{u}_{b_q,t}(\boldsymbol{\theta}))|.$$

Making use of the inequality $|xz - \tilde{x}\tilde{z}| \leq |x - \tilde{x}||z| + |x - \tilde{x}||z - \tilde{z}| + |z - \tilde{z}||x|$, we find that

$$\begin{aligned} |l_{b_q,t}(\boldsymbol{\theta}) - \tilde{l}_{b_q,t}(\boldsymbol{\theta})| \leq & |\sigma^{-1}|[|e_{x,t}(\boldsymbol{\theta}) - \tilde{e}_{x,t}(\boldsymbol{\theta})||u_{b_q,t}(\boldsymbol{\theta})| \\ & + |e_{x,t}(\boldsymbol{\theta}) - \tilde{e}_{x,t}(\boldsymbol{\theta})||u_{b_q,t}(\boldsymbol{\theta}) - \tilde{u}_{b_q,t}(\boldsymbol{\theta})| \\ & + |u_{b_q,t}(\boldsymbol{\theta}) - \tilde{u}_{b_q,t}(\boldsymbol{\theta})||e_{x,t}(\boldsymbol{\theta})|] \end{aligned}$$

for all $\boldsymbol{\theta} \in \Theta_0$. The L_p-norm of the sup of the l.h.s. is bounded by

$$\begin{aligned} & C_1 \left\| \sup_{\boldsymbol{\theta} \in \Theta_0} |e_{x,t}(\boldsymbol{\theta}) - \tilde{e}_{x,t}(\boldsymbol{\theta})| \right\|_{2p} \left\| \sup_{\boldsymbol{\theta} \in \Theta_0} |u_{b_q,t}(\boldsymbol{\theta})| \right\|_{2p} \\ & + C_1 \left\| \sup_{\boldsymbol{\theta} \in \Theta_0} |e_{x,t}(\boldsymbol{\theta}) - \tilde{e}_{x,t}(\boldsymbol{\theta})| \right\|_{2p} \left\| \sup_{\boldsymbol{\theta} \in \Theta_0} |u_{b_q,t}(\boldsymbol{\theta}) - \tilde{u}_{b_q,t}(\boldsymbol{\theta})| \right\|_{2p} \\ & + C_1 \left\| \sup_{\boldsymbol{\theta} \in \Theta_0} |u_{b_q,t}(\boldsymbol{\theta}) - \tilde{u}_{b_q,t}(\boldsymbol{\theta})| \right\|_{2p} \left\| \sup_{\boldsymbol{\theta} \in \Theta_0} |e_{x,t}(\boldsymbol{\theta})| \right\|_{2p}, \end{aligned}$$

which follows from Minkowski inequality and Cauchy-Schwartz inequality and the fact that Θ_0 is compact, so $|\sigma^{-1}| < C_1$ for some constant $C_1 < \infty$. Making use of Lemmas C2 and C3, we know that there is a positive constant

p small enough, s.t. the r.h.s. is bounded by

$$\begin{aligned}&C_2\frac{\rho^{T-t+1}}{1-\rho}+C_3\frac{\rho^{T-t+1}}{1-\rho}(T-t+1)\frac{\rho^{T-t+1}}{1-\rho}+C_4(T-t+1)\frac{\rho^{T-t+1}}{1-\rho}\\=&\frac{\rho^{T-t+1}}{1-\rho}\left(C_2+(C_3+C_4)(T-t+1)\right),\end{aligned}$$

so that

$$\left\|\sup_{\boldsymbol{\theta}\in\Theta_0}|l_{b_q,t}(\boldsymbol{\theta})-\tilde{l}_{b_q,t}(\boldsymbol{\theta})|\right\|_p\to 0\quad as\ T\to\infty.$$

Consider next a sequence $l_t^\bullet=\sup_{\boldsymbol{\theta}\in\Theta_0}|l_{b_q,T-t+1}(\boldsymbol{\theta})-\tilde{l}_{b_q,T-t+1}(\boldsymbol{\theta})|$ for $t=1,\dots,T$, and note that $\sum_{t=1}^T l_t^\bullet=\sum_{t=1}^T\sup_{\boldsymbol{\theta}\in\Theta_0}|l_{b_q,t}(\boldsymbol{\theta})-\tilde{l}_{b_q,t}(\boldsymbol{\theta})|$. We show that the l.h.s. has a finite limit as $T\to\infty$. This follows from Lemma A.2. in Meitz and Saikkonen (2011), because

$$\|\gamma^t l_t^\bullet\|_p\le\gamma^t\frac{\rho^{T-(T-t+1)+1}}{1-\rho}\left(C_2+(C_3+C_4)(T-(T-t+1))\right),$$

where γ is a positive constant s.t. $(\gamma\rho)\in(0,1)$. Because $\|\gamma^t l_t^\bullet\|_p\to 0$ as $t\to\infty$, it follows that $\lim_{T\to\infty}\sum_{t=1}^T l_t^\bullet<\infty$ a.s., and also

$$\lim_{T\to\infty}\sum_{t=1}^T\sup_{\boldsymbol{\theta}\in\Theta_0}|l_{b_q,t}(\boldsymbol{\theta})-\tilde{l}_{b_q,t}(\boldsymbol{\theta})|<\infty\ a.s.,$$

and the result follows by inequality (2.8).

Details for the proof of part *(ii)* can be found in the proof of Theorem 1 in MS13. The arguments are very similar to those given above. It is also necessary to derive results similar to those of Lemma C3 for the second order derivatives of different terms, but they do not need any additional assumptions. Details are omitted here.

Proof of Lemma D5

(i) The $m \times (P+Q+1+d)$ matrix is of the form

$$\frac{\partial}{\partial \boldsymbol{\theta}'} g_{ac,t}(\boldsymbol{\theta}) = \begin{pmatrix} \underset{m\times P}{\boldsymbol{A}(\boldsymbol{\theta})} & \underset{m\times Q}{\boldsymbol{B}(\boldsymbol{\theta})} & \boldsymbol{0}_{m\times(1+d)} \end{pmatrix},$$

with matrices $\boldsymbol{A}(\boldsymbol{\theta})$ and $\boldsymbol{B}(\boldsymbol{\theta})$ having k,l elements

$$[\boldsymbol{A}(\boldsymbol{\theta})]_{k,l} = -\frac{u_{t-l}(\boldsymbol{\theta})}{a(B)} u_{t-k}(\boldsymbol{\theta}) - u_t(\boldsymbol{\theta}) \frac{u_{t-k-l}(\boldsymbol{\theta})}{a(B)}$$

for $k = 1, \dots, m$ and $l = 1, \dots, P$,

$$[\boldsymbol{B}(\boldsymbol{\theta})]_{k,l} = \frac{u_{t+l}(\boldsymbol{\theta})}{b(B^{-1})} u_{t-k}(\boldsymbol{\theta}) + u_t(\boldsymbol{\theta}) \frac{u_{t-k+l}(\boldsymbol{\theta})}{b(B^{-1})}$$

for $k = 1, \dots, m$ and $l = 1, \dots, Q$. By Hölder's and Minkowski inequalities,

$$\begin{aligned} \left\| \sup_{\boldsymbol{\theta}\in\Theta_0} |[\boldsymbol{A}(\boldsymbol{\theta})]_{k,l}| \right\| &\le \left\| \sup_{\boldsymbol{\theta}\in\Theta_0} \left| \frac{u_{t-l}(\boldsymbol{\theta})}{a(B)} \right| \right\|_2 \left\| \sup_{\boldsymbol{\theta}\in\Theta_0} |u_{t-k}(\boldsymbol{\theta})| \right\|_2 \\ &\quad + \left\| \sup_{\boldsymbol{\theta}\in\Theta_0} |u_t(\boldsymbol{\theta})| \right\|_2 \left\| \sup_{\boldsymbol{\theta}\in\Theta_0} \left| \frac{u_{t-k-l}(\boldsymbol{\theta})}{a(B)} \right| \right\|_2 \end{aligned}$$

and

$$\begin{aligned} \left\| \sup_{\boldsymbol{\theta}\in\Theta_0} |[\boldsymbol{B}(\boldsymbol{\theta})]_{k,l}| \right\| &\le \left\| \sup_{\boldsymbol{\theta}\in\Theta_0} \left| \frac{u_{t+l}(\boldsymbol{\theta})}{b(B^{-1})} \right| \right\|_2 \left\| \sup_{\boldsymbol{\theta}\in\Theta_0} |u_{t-k}(\boldsymbol{\theta})| \right\|_2 \\ &\quad + \left\| \sup_{\boldsymbol{\theta}\in\Theta_0} |u_t(\boldsymbol{\theta})| \right\|_2 \left\| \sup_{\boldsymbol{\theta}\in\Theta_0} \left| \frac{u_{t-k+l}(\boldsymbol{\theta})}{b(B^{-1})} \right| \right\|_2, \end{aligned}$$

where the majorant sides are bounded by Lemma C2.

(ii) The $m \times (P+Q+1+d)$ derivative matrix $\frac{\partial}{\partial \boldsymbol{\theta}'} g_{hs,t}(\boldsymbol{\theta})$ can be characterized by matrix

$$\frac{\partial}{\partial \boldsymbol{\theta}'} g_{hs,t}(\boldsymbol{\theta}) = \begin{pmatrix} \underset{m\times P}{\boldsymbol{A}(\boldsymbol{\theta})} & \underset{m\times Q}{\boldsymbol{B}(\boldsymbol{\theta})} & \underset{m\times 1}{\boldsymbol{C}(\boldsymbol{\theta})} & \boldsymbol{0}_{m\times d} \end{pmatrix}.$$

The typical (k,l) element of matrix $\boldsymbol{A}(\boldsymbol{\theta})$ has a sup-norm

$$\begin{aligned}
&\left\| \sup_{\boldsymbol{\theta}\in\Theta_0} |[\boldsymbol{A}(\boldsymbol{\theta})]_{k,l}| \right\| \\
=&\left\| \sup_{\boldsymbol{\theta}\in\Theta_0} |2u_t(\boldsymbol{\theta})u_{a_l,t}(\boldsymbol{\theta})(u_{t-k}(\boldsymbol{\theta})^2-\sigma^2)+2u_{t-k}(\boldsymbol{\theta})u_{a_l,t-k}(\boldsymbol{\theta})(u_t(\boldsymbol{\theta})^2-\sigma^2)| \right\| \\
\leq& 2\left\| \sup_{\boldsymbol{\theta}\in\Theta_0} |u_t(\boldsymbol{\theta})| \right\|_4 \left\| \sup_{\boldsymbol{\theta}\in\Theta_0} |u_{a_l,t}(\boldsymbol{\theta})| \right\|_4 \left\| \sup_{\boldsymbol{\theta}\in\Theta_0} |u_{t-k}(\boldsymbol{\theta})^2-\sigma^2| \right\|_2 \\
&+2\left\| \sup_{\boldsymbol{\theta}\in\Theta_0} |u_{t-k}(\boldsymbol{\theta})| \right\|_4 \left\| \sup_{\boldsymbol{\theta}\in\Theta_0} |u_{a_l,t-k}(\boldsymbol{\theta})| \right\|_4 \left\| \sup_{\boldsymbol{\theta}\in\Theta_0} |u_t(\boldsymbol{\theta})^2-\sigma^2| \right\|_2 .
\end{aligned}$$

This is bounded by the Lemma C2, and compactness of set Θ_0. Matrix $\boldsymbol{B}(\boldsymbol{\theta})$ can be shown to have bounded sup-norm with exactly the same arguments using the point *(iii)* in Lemma C2, instead of *(ii)*.

The sup-norm of the typical k element of $\boldsymbol{C}(\boldsymbol{\theta})$ can be bounded by

$$\begin{aligned}
\left\| \sup_{\boldsymbol{\theta}\in\Theta_0} |[\boldsymbol{C}(\boldsymbol{\theta})]_k| \right\| \leq& 4\left\| \sup_{\boldsymbol{\theta}\in\Theta_0} |\sigma^3| \right\| + 2\left\| \sup_{\boldsymbol{\theta}\in\Theta_0} |u_{t-k}(\boldsymbol{\theta})^2| \right\|_2 \left\| \sup_{\boldsymbol{\theta}\in\Theta_0} |\sigma| \right\|_2 \\
&+\left\| \sup_{\boldsymbol{\theta}\in\Theta_0} |u_t(\boldsymbol{\theta})^2| \right\|_2 \left\| \sup_{\boldsymbol{\theta}\in\Theta_0} |\sigma| \right\|_2 .
\end{aligned}$$

The majorant is finite by compactness of Θ_0 and by Lemma C2.

Remaining parts of the proof of Lemma D6

In order to justify part A *(iv)*, we will find out the form of the asymptotic covariance matrix of $T^{-1/2}\sum_{t=1}^{T}\boldsymbol{\xi}_{ac,t}(\boldsymbol{\theta}_0)$, and show that it is finite and positive definite. It can be shown, that if exists, the covariance matrix is of the form

$$\boldsymbol{\Sigma}_{\boldsymbol{\xi}_{ac}}(\boldsymbol{\theta}_0) = \sum_{s=-\infty}^{\infty} \mathrm{E}[\xi_{ac,t}(\boldsymbol{\theta}_0)\xi'_{ac,t-s}(\boldsymbol{\theta}_0)].$$

For any $s \in \mathbb{Z}$, let us write

$$\begin{aligned}
\mathrm{E}[\xi_{ac,t}(\boldsymbol{\theta}_0)\xi'_{ac,t-s}(\boldsymbol{\theta}_0)] &= \begin{pmatrix} \mathrm{E}[l_{\boldsymbol{\theta},t}(\boldsymbol{\theta}_0)l'_{\boldsymbol{\theta},t-s}(\boldsymbol{\theta}_0)] & \mathrm{E}[l_{\boldsymbol{\theta},t}(\boldsymbol{\theta}_0)g'_{ac,t-s}(\boldsymbol{\theta}_0)] \\ \mathrm{E}[\boldsymbol{g}_{ac,t}(\boldsymbol{\theta}_0)l'_{\boldsymbol{\theta},t-s}(\boldsymbol{\theta}_0)] & \mathrm{E}[\boldsymbol{g}_{ac,t}(\boldsymbol{\theta}_0)g'_{ac,t-s}(\boldsymbol{\theta}_0)] \end{pmatrix} \\
&\overset{def.}{=} \begin{pmatrix} \mathrm{E}[l_{\boldsymbol{\theta},t}(\boldsymbol{\theta}_0)l'_{\boldsymbol{\theta},t-s}(\boldsymbol{\theta}_0)] & \Psi^{s\prime}_{ac} \\ \Psi^{s}_{ac} & \mathcal{H}^{s}_{ac} \end{pmatrix}.
\end{aligned}$$

The first diagonal element of $\boldsymbol{\Sigma}_{\xi_{ac}}$ is $\boldsymbol{\mathcal{I}}$ given in (2.7). The second block of matrix Ψ^s_{ac} has a representative element

$$\begin{aligned}
[\Psi^s_{ac,B}]_{k,l} &= \mathrm{E}\left[\varepsilon_t\varepsilon_{t-k}e_{x,t-s}\frac{1}{\sigma_0}\sum_{j=0}^{\infty}\psi^{(b)}_{0,j}\varepsilon_{t-s+l+j}\right] \\
&= \begin{cases} 0, & k \neq s \\ -\sigma_0^2\psi^b_{0,k-l} & k = s \end{cases}.
\end{aligned}$$

The third block of Ψ^s_{ac} has a representative element

$$[\Psi^s_C]_k = \mathrm{E}\left[-\frac{1}{\sigma_0}\varepsilon_t\varepsilon_{t-k}\left(e_{x,t-s}\frac{\varepsilon_{t-s}}{\sigma_0^2}+1\right)\right] = 0$$

for all s. This is evident from observing that k is always a positive integer and then using Lemma C2.

The last block, $\Psi^s_{ac,D}$, has a representative element $[\Psi^s_D]_{k,l} = \mathrm{E}[\varepsilon_t\varepsilon_{t-k}e_{\lambda_l,t-s}] = 0$, since $k > 0$ and Lemma C2.

The off-diagonal block of the covariance matrix $\boldsymbol{\Sigma}_{\xi_{ac}}$ is,

$$\sum_{s=-\infty}^{\infty}\boldsymbol{\Psi}_{ac} = \sigma_0^2\begin{pmatrix}\boldsymbol{\Lambda}_m & \mathbf{0}_{m\times(1+d)}\end{pmatrix},$$

$\boldsymbol{\Lambda}_m$ given in (2.5).

Next, we take a look at the lower diagonal element $\mathcal{H}^s_{ac}(\boldsymbol{\theta}_0) = \mathrm{E}[\boldsymbol{g}_{ac,t}(\boldsymbol{\theta}_0)\boldsymbol{g}'_{ac,t-s}(\boldsymbol{\theta}_0)]$. This matrix has a representative (k,l) element $\mathrm{E}[\varepsilon_t\varepsilon_{t-k}\varepsilon_{t-s}\varepsilon_{t-s-l}]$, so the lower diagonal block of matrix $\boldsymbol{\Sigma}_{\xi_{ac}}$ is thus $\sigma_0^4\boldsymbol{I}_{m\times m}$.

In order to conclude that the obtained covariance matrix is positive definite, we claim that there is a $(P+Q+1+d+m)$ vector $\mathcal{X}$ s.t. $\mathrm{Cov}(\mathcal{X}) =$

$Y_{ac}(\boldsymbol{\theta}_0)$, where

$$\mathcal{X} = \begin{pmatrix} x_{a,t} \\ x_{b,t} \\ \left(e_{x,t}\dfrac{\varepsilon_t}{\sigma_0} - 1\right)\dfrac{1}{\sigma_0} \\ e_{\lambda,t} \\ g_{ac,t} \end{pmatrix},$$

$$x_{a,t} = \begin{pmatrix} -\dfrac{1}{\sigma_0}\sum_{j=0}^{\infty}\psi_{0,j}^{(a)}\zeta_{t,j+1}^{(a)} \\ \vdots \\ -\dfrac{1}{\sigma_0}\sum_{j=0}^{\infty}\psi_{0,j}^{(a)}\zeta_{t,j+P}^{(s)} \end{pmatrix}, \quad x_{b,t} = \begin{pmatrix} \dfrac{1}{\sigma_0}\sum_{j=0}^{\infty}\psi_{0,j}^{(b)}\zeta_{t,j+1}^{(b)} \\ \vdots \\ \dfrac{1}{\sigma_0}\sum_{j=0}^{\infty}\psi_{0,j}^{(b)}\zeta_{t,j+Q}^{(b)} \end{pmatrix},$$

$\zeta_{t,j+p}^{(a)} = e_{x,t}\varepsilon_{t-j-p}$, and $\zeta_{t,j+q}^{(b)} e_{x,t-j-q}\varepsilon_t$. Note that random variables $\zeta_{t,j}^{(a)}$ and $\zeta_{t,k}^{(a)}$ are correlated only for $j = k$ and the same applies for the correlations of the pairs $(\zeta_{t,j}^{(b)}, \zeta_{t,k}^{(b)})$ and $(\zeta_{t,j}^{(a)}, \zeta_{t,k}^{(b)})$. Using this fact we can further write the vectors $x_{a,t}$ and $x_{b,t}$ as a sum of two uncorrelated random vectors as $x_{a,t} = x_{a,t}^{(1)} + x_{a,t}^{(2)}$ and $x_{b,t} = x_{b,t}^{(1)} + x_{b,t}^{(2)}$. Let us define $K = \max\{P, Q\}$, and write

$$x_{a,t}^{(1)} = \begin{pmatrix} -\dfrac{1}{\sigma_0}\sum_{j=0}^{K-1}\psi_{0,j}^{(a)}\zeta_{t,j+1}^{(a)} \\ -\dfrac{1}{\sigma_0}\sum_{j=0}^{K-2}\psi_{0,j}^{(a)}\zeta_{t,j+2}^{(a)} \\ \vdots \\ -\dfrac{1}{\sigma_0}\sum_{j=0}^{K-P}\psi_{0,j}^{(a)}\zeta_{t,j+P}^{(a)} \end{pmatrix}, \quad x_{a,t}^{(2)} = \begin{pmatrix} -\dfrac{1}{\sigma_0}\sum_{j=K-1+1}^{\infty}\psi_{0,j}^{(a)}\zeta_{t,j+1}^{(a)} \\ -\dfrac{1}{\sigma_0}\sum_{j=K-2+1}^{\infty}\psi_{0,j}^{(a)}\zeta_{t,j+2}^{(a)} \\ \vdots \\ -\dfrac{1}{\sigma_0}\sum_{j=K-P+1}^{\infty}\psi_{0,j}^{(a)}\zeta_{t,j+P}^{(a)} \end{pmatrix},$$

$$x_{b,t}^{(1)} = \begin{pmatrix} \frac{1}{\sigma_0}\sum_{j=0}^{K-1}\psi_{0,j}^{(b)}\zeta_{t,j+1}^{(b)} \\ \frac{1}{\sigma_0}\sum_{j=0}^{K-2}\psi_{0,j}^{(b)}\zeta_{t,j+2}^{(b)} \\ \vdots \\ \frac{1}{\sigma_0}\sum_{j=0}^{K-Q}\psi_{0,j}^{(b)}\zeta_{t,j+Q}^{(b)} \end{pmatrix} \quad \text{and} \quad x_{b,t}^{(2)} = \begin{pmatrix} \frac{1}{\sigma_0}\sum_{j=K-1+1}^{\infty}\psi_{0,j}^{(b)}\zeta_{t,j+1}^{(b)} \\ \frac{1}{\sigma_0}\sum_{j=K-2+1}^{\infty}\psi_{0,j}^{(b)}\zeta_{t,j+2}^{(b)} \\ \vdots \\ \frac{1}{\sigma_0}\sum_{j=K-Q+1}^{\infty}\psi_{0,j}^{(b)}\zeta_{t,j+Q}^{(b)} \end{pmatrix}.$$

Now we can split $\mathcal{X}$ into two uncorrelated parts as $\mathcal{X} = \mathcal{X}_1 + \mathcal{X}_2$ with obvious definitions

$$\mathcal{X}_1 = \begin{pmatrix} x_{a,t}^{(1)} \\ x_{b,t}^{(1)} \\ \frac{1}{\sigma_0^2}\varepsilon_t e_{x,t} \\ e_{\lambda,t} \\ \boldsymbol{g}_{ac,t}(\boldsymbol{\theta}_0) \end{pmatrix} \quad \text{and} \quad \mathcal{X}_2 = \begin{pmatrix} x_{a,t}^{(2)} \\ x_{b,t}^{(2)} \\ -\frac{1}{\sigma_0} \\ 0 \\ 0 \end{pmatrix}.$$

In order to show that $\mathrm{Cov}(\mathcal{X})$ is positive definite, it suffices to show that $\mathrm{Cov}(\mathcal{X}_1)$ is positive definite. To this end, let us define vectors $z_t^{(a)} = (\zeta_{t,1}^{(a)}, \ldots, \zeta_{t,K}^{(a)})$ and $z_t^{(b)} = (\zeta_{t,1}^{(b)}, \ldots, \zeta_{t,K}^{(b)})$ and write the random vector $\mathcal{X}_1$ yet in a different form as

$$\mathcal{X}_1 = \begin{pmatrix} \phi^a & 0 & 0 & 0 & 0 \\ 0 & \phi^{(b)} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\sigma_0^2} & 0 & 0 \\ 0 & 0 & 0 & I_{d\times d} & 0 \\ 0 & 0 & 0 & 0 & I_{m\times m} \end{pmatrix} \begin{pmatrix} z_t^{(a)} \\ z_t^{(b)} \\ \zeta_{t,0}^{(a)} \\ e_{\lambda,t} \\ \boldsymbol{g}_{ac,t}(\boldsymbol{\theta}_0) \end{pmatrix}$$

with $(P \times K)$ and $(Q \times K)$ matrices

$$\phi^{(a)} = \begin{pmatrix} -\frac{1}{\sigma_0}\psi^{(a)}_{0,0} & -\frac{1}{\sigma_0}\psi^{(a)}_{0,1} & -\frac{1}{\sigma_0}\psi^{(a)}_{0,2} & \cdots & -\frac{1}{\sigma_0}\psi^{(a)}_{0,K-1} \\ 0 & -\frac{1}{\sigma_0}\psi^{(a)}_{0,0} & -\frac{1}{\sigma_0}\psi^{(a)}_{0,1} & \cdots & -\frac{1}{\sigma_0}\psi^{(a)}_{0,K-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & -\frac{1}{\sigma_0}\psi^{(a)}_{0,K-P} \end{pmatrix} \quad \text{and}$$

$$\phi^{(b)} = \begin{pmatrix} \frac{1}{\sigma_0}\psi^{(b)}_{0,0} & \frac{1}{\sigma_0}\psi^{(b)}_{0,1} & \frac{1}{\sigma_0}\psi^{(b)}_{0,2} & \cdots & \frac{1}{\sigma_0}\psi^{(b)}_{0,K-1} \\ 0 & \frac{1}{\sigma_0}\psi^{(b)}_{0,0} & \frac{1}{\sigma_0}\psi^{(b)}_{0,1} & \cdots & \frac{1}{\sigma_0}\psi^{(b)}_{0,K-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{1}{\sigma_0}\psi^{(b)}_{0,K-Q} \end{pmatrix}.$$

These matrices have ranks P and Q respectively, so the first matrix in the definition of $\mathcal{X}_1$ has a full row rank of $(P+Q+1+d+m)$. This implies that $\mathrm{Cov}(\mathcal{X}_1)$ is positive definite if and only if vector

$$(z_t^{(a)}, z_t^{(b)}, \zeta_{t,0}^{(a)}, e_{\lambda,t}, \boldsymbol{g}_{ac,t}(\boldsymbol{\theta}_0)) \tag{2.9}$$

has a positive definite covariance matrix. To this end, let $K < m$. The opposite $K \geq m$ can be shown in a similar manner.

If $K < m$, let us re-organize the vector (2.9) as

$$\left(\zeta_{t,1}^{(a)}, \zeta_{t,1}^{(b)}, \varepsilon_t\varepsilon_{t-1}, \ldots, \zeta_{t,K}^{(a)}, \zeta_{t,K}^{(b)}, \varepsilon_t\varepsilon_{t-K}, \varepsilon_t\varepsilon_{t-K-1}, \ldots, \varepsilon_t\varepsilon_{t-m}, \zeta_{t,0}^{(a)}, e_{\lambda,t}\right).$$

This vector has a block diagonal covariance matrix

$$\begin{pmatrix} \varphi_1 & 0 & \dots & 0 & 0 & 0 \\ 0 & \varphi_2 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \varphi_K & 0 & 0 \\ 0 & 0 & \dots & 0 & \Sigma & 0 \\ 0 & 0 & \dots & 0 & 0 & \vartheta \end{pmatrix},$$

where $\varphi_1, \dots, \varphi_K$ are 3×3 matrices,

$$\varphi_k = \mathrm{Cov}\begin{pmatrix} \zeta_{t,k}^{(a)} \\ \zeta_{t,k}^{(b)} \\ \varepsilon_t \varepsilon_{t-k} \end{pmatrix} \quad \text{for } k = 1, \dots, K,$$

$$\Sigma = \mathrm{Cov}\begin{pmatrix} \varepsilon_t \varepsilon_{t-K-1} \\ \vdots \\ \varepsilon_t \varepsilon_{t-m} \end{pmatrix} \quad \text{and} \quad \vartheta = \mathrm{Cov}\begin{pmatrix} \zeta_{t,0}^{(a)} \\ e_{\lambda,t} \end{pmatrix}.$$

This covariance matrix is positive definite because all the blocks are positive definite matrices. To show this, we start from the last one. Matrix ϑ is positive definite if $a_1 \zeta_{t,0}^{(a)} + a_2' e_{\lambda,t} = 0$ only for $a_1 = 0$ and a_2 is zero vector. Taking expectations and using Lemma C2 yields $a_1 \sigma_0 = 0$ that holds only for $a_1 = 0$. If $a_1 = 0$, it must be that $a_2' e_{\lambda,t} = 0$, but by Assumption 3, A.1. (iii), this holds only if a_2 is a zero vector. That is, ϑ is positive definite.

The second covariance matrix Σ is positive definite as it is $(m-K) \times (m-K)$ diagonal matrix with σ_0^4 on the diagonal.

In order to show that φ_k is positive definite for all $k = 1, \dots, K$ we have to show that

$$a_1 \zeta_{t,k}^{(a)} + a_2 \zeta_{t,k}^{(b)} + a_3 \varepsilon_t \varepsilon_{t-k} = 0 \tag{2.10}$$

only if $a_1 = a_2 = a_3 = 0$. To this end, we multiply both sides of (2.10) by

$\zeta_{t,k}^{(a)}$ and take expectations of both sides, yielding $a_3\sigma_0^3 = a_1\sigma_0^2\mathrm{E}[e_{x,t}^2] + a_2\sigma_0^2$. Multiplying both sides of (2.10) by $\zeta_{t,k}^{(b)}$ and taking expectations, yields

$$a_3\sigma_0^3 = a_1\sigma_0^2 + a_2\sigma_0^2\mathrm{E}[e_{x,t}^2]. \tag{2.11}$$

Equating the r.h.s.'s of both equations yields $a_1\sigma_0^2(1+\mathrm{E}[e_{x,t}^2]) = a_2\sigma_0^2(1+\mathrm{E}[e_{x,t}^2]1)$. As we have assumed non-Gaussian distribution for ε_t, terms in the parenthesis are always greater than 2 (Andrews et al. (2006), Remark 2), we obtain $a_1 = a_2$. If $a_1 = a_2 = 0$, squaring both sides of (2.10) and taking expectation conditional on $\mathcal{F}_t$ would give $a_3^2\varepsilon_t^2\varepsilon_{t-k}^2 = 0$, which would be contradictory.

Assuming that $a_1 = a_2 = a \neq 0$ and solving for a_3 in (2.11) gives

$$a_3 = \frac{a(1+\mathrm{E}[e_{x,t}^2])}{\sigma_0}.$$

Multiplying both sides of (2.10) by $\varepsilon_t\varepsilon_{t-k}$ and taking expectations gives $a_3 = 2a\sigma_0^{-1}$. But this contradicts with (2.11), because together these expressions for a_3 would imply that $2 = 1+\mathrm{E}[e_{x,t}^2]$, but as we have already noted, this cannot hold because $1+\mathrm{E}[e_{x,t}^2] > 2$. That is, (2.10) holds only for $a_1 = a_2 = a_3 = 0$ and thus φ_k is a positive definite matrix.

These arguments imply that the covariance matrix Υ_{ac} is positive definite and Lemma D6 A (iv) holds.

Proof of Lemma D6 B (ii) is simpler than the previous proof, because subvectors of $\xi_{hs,t}(\boldsymbol{\theta}_0)$, $l_t(\boldsymbol{\theta}_0)$ and $g_{hs,t-s}(\boldsymbol{\theta}_0)$ are not correlated on any lag length s. Let us recall the notation

$$\mathrm{E}[\xi_{hs,t}(\boldsymbol{\theta}_0)\xi_{hs,t-s}(\boldsymbol{\theta}_0)'] = \begin{pmatrix} \mathrm{E}[l_{\boldsymbol{\theta},t}(\boldsymbol{\theta}_0)l_{\boldsymbol{\theta},t-s}(\boldsymbol{\theta}_0)'] & \Psi_{hs}^{s\prime} \\ \Psi_{hs}^{s} & \mathcal{H}_{hs}^{s} \end{pmatrix}$$

and

$$\boldsymbol{\Sigma}_{\xi_{hs}} = \sum_{s=-\infty}^{\infty} \mathrm{E}\left[\xi_{hs,t}(\boldsymbol{\theta}_0)\xi_{hs,t-s}(\boldsymbol{\theta})'\right] = \begin{pmatrix} \mathcal{I} & \Upsilon_{hs,\Psi}' \\ \Upsilon_{hs,\Psi} & \Upsilon_{hs,\mathcal{H}} \end{pmatrix}.$$

Let us denote $\kappa_t \overset{def}{=} \varepsilon_t^2 - \sigma_0^2$ and note that κ_t is an iid sequence with mean zero and $\mathrm{E}[\kappa_t^2] = \mathrm{E}[\varepsilon_t^4] - \sigma_0^4$. Moreover, $\mathrm{E}[\kappa_t e_{x,t}] = 0$ by points (vi) and $(viii)$, $\mathrm{E}[\kappa_t e_{x,t} \varepsilon_t] = -2\sigma_0^3 < \infty$ by points (vii) and $(viii)$ and $\mathrm{E}[\kappa_t e_{\lambda_l,t}] = 0$ by (x) and (xi) of Lemma C2.

For $s = 0$, matrix $\mathcal{H}_{hs}^s(\boldsymbol{\theta}_0)$ has a typical k, l element

$$\mathrm{E}[\kappa_t^2 \kappa_{t-k} \kappa_{t-l}] = \begin{cases} \mathrm{E}[\kappa_t^2]^2 & k = l \\ 0 & k \neq l \end{cases} .$$

For $s \neq 0$ we have the expression $\mathcal{H}_{hs}^s = \mathrm{E}[\kappa_t \kappa_{t-k} \kappa_{t-s} \kappa_{t-s-l}] = 0$, because if $s = k$, then $t - s - l = t - k - l \neq t$ because k and l are positive. There is no way to have an expression without odd powers of κ_t. Independence of κ_t and κ_{t-s} for $s \neq 0$ implies the zero expectation. To sum up, $\boldsymbol{\Sigma}_{\xi_{hs}}$ has a clearly positive definite lower diagonal block $\mathrm{E}[\kappa_t^2]^2 \boldsymbol{I}_{m \times m}$.

The summands in the off-diagonal block of Υ_{hs} can be written as

$$\Psi_{hs}^s(\boldsymbol{\theta}_0) = \begin{pmatrix} \underset{m \times P}{\Psi_{hs,A}^s} & \underset{m \times Q}{\Psi_{hs,B}^s} & \underset{m \times 1}{\Psi_{hs,C}^s} & \underset{m \times d}{\Psi_{hs,D}^s} \end{pmatrix}.$$

The first block is $(m \times P)$ matrix and it has a typical element

$$[\Psi_{hs,A}^s]_{k,l} = \mathrm{E}\left[-\kappa_t \kappa_{t-k} e_{x,t-s} \sum_{j=0}^{\infty} \psi_{0,j}^{(a)} \varepsilon_{t-s-l-j} \right] = 0 \quad \forall s.$$

For $s = 0$ it can be seen by writing the expectation as

$$-\mathrm{E}\left[\kappa_t e_{x,t}\right] \mathrm{E}\left[\kappa_{t-k} \sum_{j=0}^{\infty} \psi_{0,j}^{(a)} \varepsilon_{t-l-j} \right] = 0,$$

and noting that the first term has finite expectation and the latter is zero. For $s \neq 0$, expectation can always be written as

$$-\mathrm{E}\left[\kappa_t\right] \mathrm{E}\left[\kappa_{t-k} e_{x,t-s} \sum_{j=0}^{\infty} \psi_{0,j}^{(a)} \varepsilon_{t-s-l-j} \right] = 0.$$

The first expectation is clearly zero, and the latter is as well, because for $s = k$, we have $t - k - l - j < t - k$ for positive j, l and k.

The second block has k, l element

$$[\Psi^s_{hs,B}(\boldsymbol{\theta}_0)\ ]_{k,l} = \mathrm{E}\left[\kappa_t \kappa_{t-k} e_{x,t-s} \sum_{j=0}^{\infty} \psi^{(b)}_{0,j} \varepsilon_{t-s+l+j}\right] = 0 \quad \forall s.$$

If $s = 0$, the expectation is

$$\mathrm{E}[\kappa_t e_{x,t}]\mathrm{E}[\kappa_{t-k}]\mathrm{E}\left[\sum_{j=0}^{\infty} \psi^{(b)}_{0,j} \varepsilon_{t+l+j}\right] = 0,$$

because the last two terms are zero. For $s = k$ the expectation is

$$\mathrm{E}[\kappa_{t-k} e_{x,t-k}]\mathrm{E}[\kappa_t \sum_{j=0}^{\infty} \psi^{(b)}_{0,j} \varepsilon_{t-s+l+j}] = 0,$$

because the latter term has zero expectation for all s.

The third block has a representative element

$$[\Psi^s_{hs,C}]_k = \mathrm{E}\left[-\frac{1}{\sigma_0}\kappa_t \kappa_{t-k}\left(e_{x,t-s}\frac{\varepsilon_{t-s}}{\sigma_0^2} + 1\right)\right] = 0 \quad \forall\ s.$$

There is no chance of having coinciding indices for κ_t. If $s = k$, simply note that $\mathrm{E}[\kappa_t] = 0$. Otherwise, note that $\mathrm{E}[\kappa_t \kappa_{t-k}] = 0$.

The k, l element of the last block can be written as

$$[\Psi^s_{hs,D}]_{k,l} = \mathrm{E}\left[\kappa_t \kappa_{t-k} e_{\lambda_l,t-s}\right] = 0 \quad \forall\ s,$$

because if $s = 0$ or $s = k$, the expectation is $\mathrm{E}[\kappa_t e_{\lambda_l,t}]\mathrm{E}[\kappa_{t-k}] = 0$. Otherwise it is $\mathrm{E}[\kappa_t]\mathrm{E}[\kappa_{t-k}]\mathrm{E}[e_{\lambda_l,t-s}] = 0$.

To sum up, all the blocks are zero matrices for all s:

$$\Upsilon_{hs,\Psi} = \mathbf{0}_{m\times(P+D+1+d)}.$$

The first diagonal block of the asymptotic covariance matrix was positive definite $\mathcal{I}$. The block diagonal structure of $\boldsymbol{\Sigma}_{\xi_{hs}}$, with positive definite blocks on the diagonal, implies the result.

Proof of Lemma D7

In order to prove D7 (i), note that

$$\sup_{\boldsymbol{\theta}\in\Theta_0}\left|T^{-1/2}\sum_{t=1}^{T}g_{ac,t}(\boldsymbol{\theta})-T^{-1/2}\sum_{t=1}^{T}\tilde{g}_{ac,t}(\boldsymbol{\theta})\right|\leq T^{-1/2}\sum_{t=1}^{T}\sup_{\boldsymbol{\theta}\in\Theta_0}\left|g_{ac,t}(\boldsymbol{\theta})-\tilde{g}_{ac,t}(\boldsymbol{\theta})\right|.$$

The result follows after showing that the sum on the r.h.s. has a finite limit. To this end, write the k^{th} element of $\sup_{\boldsymbol{\theta}\in\Theta_0}|g_t(\boldsymbol{\theta})-\tilde{g}_t(\boldsymbol{\theta})|$ as

$$\sup_{\boldsymbol{\theta}\in\Theta_0}|u_t(\boldsymbol{\theta})u_{t-k}(\boldsymbol{\theta})-\tilde{u}_t(\boldsymbol{\theta})\tilde{u}_{t-k}(\boldsymbol{\theta})|$$

and bound it from above using inequality

$$|xy-\tilde{x}\tilde{y}|\leq|x-\tilde{x}|\,|y|+|x-\tilde{x}|\,|y-\tilde{y}|+|y-\tilde{y}|\,|x| \tag{2.12}$$

by

$$\sup_{\boldsymbol{\theta}\in\Theta_0}|u_t(\boldsymbol{\theta})-\tilde{u}_t(\boldsymbol{\theta})|\sup_{\boldsymbol{\theta}\in\Theta_0}|u_{t-k}(\boldsymbol{\theta})|+\sup_{\boldsymbol{\theta}\in\Theta_0}|u_t(\boldsymbol{\theta})-\tilde{u}_t(\boldsymbol{\theta})|\sup_{\boldsymbol{\theta}\in\Theta_0}|u_{t-k}(\boldsymbol{\theta})-\tilde{u}_{t-k}(\boldsymbol{\theta})|$$

$$+\sup_{\boldsymbol{\theta}\in\Theta_0}|u_{t-k}(\boldsymbol{\theta})-\tilde{u}_{t-k}(\boldsymbol{\theta})|\sup_{\boldsymbol{\theta}\in\Theta_0}|u_{t-k}(\boldsymbol{\theta})|.$$

Consider the L_1-norm and use Minkowsis' inequality and Cauchy-Schwartz inequality to bound it by

$$\begin{aligned}
&\left\|\sup_{\boldsymbol{\theta}\in\Theta_0}|u_t(\boldsymbol{\theta})-\tilde{u}_t(\boldsymbol{\theta})|\right\|_2\left\|\sup_{\boldsymbol{\theta}\in\Theta_0}|u_{t-k}(\boldsymbol{\theta})|\right\|_2\\
&+\left\|\sup_{\boldsymbol{\theta}\in\Theta_0}|u_t(\boldsymbol{\theta})-\tilde{u}_t(\boldsymbol{\theta})|\right\|_2\left\|\sup_{\boldsymbol{\theta}\in\Theta_0}|u_{t-k}(\boldsymbol{\theta})-\tilde{u}_{t-k}(\boldsymbol{\theta})|\right\|_2\\
&+\left\|\sup_{\boldsymbol{\theta}\in\Theta_0}|u_{t-k}(\boldsymbol{\theta})-\tilde{u}_{t-k}(\boldsymbol{\theta})|\right\|_2\left\|\sup_{\boldsymbol{\theta}\in\Theta_0}|u_{t_k}(\boldsymbol{\theta})|\right\|_2\\
\leq&C_1\rho^{T-t+1}+C_2\rho^{T-t+1}\rho^{T-t+k+1}+C_3\rho^{T-t+1},
\end{aligned}\tag{2.13}$$

where the last inequality follows from Lemmas C2 and C3. That is, $\sup_{\boldsymbol{\theta}\in\Theta_0} |g_{ac,t}(\boldsymbol{\theta}) - \tilde{g}_{ac,t}(\boldsymbol{\theta})|$ converges to zero in L_1-norm as $T \to \infty$.

Denote $g_t^{\bullet} = \sup_{\boldsymbol{\theta}\in\Theta_0} |g_{ac,T-t+1}(\boldsymbol{\theta}) - \tilde{g}_{ac,T-t+1}(\boldsymbol{\theta})|$, and note

$$\sum_{t=1}^{T} g_t^{\bullet} = \sum_{t=1}^{T} \sup_{\boldsymbol{\theta}\in\Theta_0} |g_{ac,t}(\boldsymbol{\theta}) - \tilde{g}_{ac,t}(\boldsymbol{\theta})|\,.$$

The wanted result follows if the l.h.s. has a finite limit, which is the case, since using inequality (2.13),

$$\left\|\gamma^t g_t^{\bullet}\right\| \leq (\gamma\rho)^t \rho^{-1}(C_1 + C_2\rho^{t+k} + C_3\rho^k) \to 0 \quad \text{as } t \to \infty$$

for some constant γ s.t. $|\gamma| > 1$ and $|\gamma\rho| < 1$. By Lemma A.2. in Meitz and Saikkonen (2011), $\lim_{T\to\infty} \sum_{t=1}^{T} g_t^{\bullet} < \infty$ a.s. These arguments hold for each element in vector $\sup_{\boldsymbol{\theta}\in\Theta_0} |g_{ac,t}(\boldsymbol{\theta}) - \tilde{g}_{ac,t}(\boldsymbol{\theta})|$ so the result follows elementwise.

For part (ii), denote $\kappa_t(\boldsymbol{\theta}) \overset{def}{=} u_t(\boldsymbol{\theta})^2 - \sigma^2$ and $\tilde{\kappa}_t(\boldsymbol{\theta}) \overset{def}{=} \tilde{u}_t(\boldsymbol{\theta})^2 - \sigma^2$. Then

$$\begin{aligned}
&\left\|\sup_{\boldsymbol{\theta}\in\Theta_0} |\kappa_t(\boldsymbol{\theta}) - \tilde{\kappa}_t(\boldsymbol{\theta})|\right\|_2 \\
\leq &\left\|\sup_{\boldsymbol{\theta}\in\Theta_0} |u_t(\boldsymbol{\theta}) - \tilde{u}_t(\boldsymbol{\theta})|\right\|_4 \left[\left\|\sup_{\boldsymbol{\theta}\in\Theta_0} |u_t(\boldsymbol{\theta})|\right\|_4 + \left\|\sup_{\boldsymbol{\theta}\in\Theta_0} |\tilde{u}_t(\boldsymbol{\theta})|\right\|_4\right] \\
\leq &C\rho^{T-t+1}
\end{aligned} \tag{2.14}$$

for some $C < \infty$ and $|\rho| < 1$ by Lemma C2, C3, and the fact that

$$\left\|\sup_{\boldsymbol{\theta}\in\Theta_0} |\tilde{u}_t(\boldsymbol{\theta})|\right\|_4 \leq \sum_{j=0}^{T-t} \sup_{\boldsymbol{\theta}\in\Theta_0} |\psi_j^{(b)}| \left\|\sup_{\boldsymbol{\theta}\in\Theta_0} \left|a(B)y_{t+j}\right|\right\|_4 < \infty,$$

because Θ_0 is compact and $\mathrm{E}[y_t^4] < \infty$. Moreover, $\|\sup_{\boldsymbol{\theta}\in\Theta_0} |\kappa_t(\boldsymbol{\theta})|\|_2 < \infty$ because of Lemma C2 (i) and compactness of Θ_0.

Following the lines of the proof of part (i), we show that the sum

$$\sum_{t=1}^{T} \sup_{\boldsymbol{\theta}\in\Theta_0} \left|g_{hs,t}(\boldsymbol{\theta}) - \tilde{g}_{hs,t}(\boldsymbol{\theta})\right|$$

converges by showing that all the elements converges to zero in L_1-norm as $T \to \infty$. To see this, consider the k^{th} element and write

$$
\begin{aligned}
&\left\| \sup_{\boldsymbol{\theta}\in\Theta_0} |\kappa_t(\boldsymbol{\theta})\kappa_{t-k}(\boldsymbol{\theta}) - \tilde{\kappa}_t(\boldsymbol{\theta})\tilde{\kappa}_{t-k}(\boldsymbol{\theta})| \right\| \\
\leq & \left\| \sup_{\boldsymbol{\theta}\in\Theta_0} |\kappa_t(\boldsymbol{\theta}) - \tilde{\kappa}_t(\boldsymbol{\theta})| \right\|_2 \left\| \sup_{\boldsymbol{\theta}\in\Theta_0} |\kappa_{t-k}(\boldsymbol{\theta})| \right\|_2 \\
& + \left\| \sup_{\boldsymbol{\theta}\in\Theta_0} |\kappa_t(\boldsymbol{\theta}) - \tilde{\kappa}_t(\boldsymbol{\theta})| \right\|_2 \left\| \sup_{\boldsymbol{\theta}\in\Theta_0} |\kappa_{t-k}(\boldsymbol{\theta}) - \tilde{\kappa}_{t-k}(\boldsymbol{\theta})| \right\|_2 \\
& + \left\| \sup_{\boldsymbol{\theta}\in\Theta_0} |\kappa_{t-k}(\boldsymbol{\theta}) - \tilde{\kappa}_{t-k}(\boldsymbol{\theta})| \right\|_2 \left\| \sup_{\boldsymbol{\theta}\in\Theta_0} |\kappa_t(\boldsymbol{\theta})| \right\|_2 \\
\leq & C_1\rho^{T-t+1} + C_2\rho^{2(T-t+1)+k} + C_3\rho^{T-t+k+1}.
\end{aligned}
$$

The majorant clearly goes to zero as $T \to \infty$. The rest follows exactly as in the previous part. The same arguments also apply for all the elements in the vector, so the statement is true elementwise.

Part (iii) follows by similar arguments. Start by noting that

$$
\begin{aligned}
&\sup_{\boldsymbol{\theta}\in\Theta_0} \left| T^{-1}\sum_{t=1}^{T} g_{ac,\boldsymbol{\theta},t}(\boldsymbol{\theta}) - T^{-1}\sum_{t=1}^{T} \tilde{g}_{ac,\boldsymbol{\theta},t}(\boldsymbol{\theta}) \right| \\
\leq & T^{-1}\sum_{t=1}^{T} \sup_{\boldsymbol{\theta}\in\Theta_0} |g_{ac,\boldsymbol{\theta},t}(\boldsymbol{\theta}) - \tilde{g}_{ac,\boldsymbol{\theta},t}(\boldsymbol{\theta})|.
\end{aligned}
\tag{2.15}
$$

The Sum on the r.h.s. has a finite limit as $T \to \infty$. To see this, let us consider the derivatives w.r.t. b_q. Write the derivative of the k^{th} element of $\sup_{\boldsymbol{\theta}\in\Theta_0} |g_{ac,\boldsymbol{\theta},t}(\boldsymbol{\theta}) - \tilde{g}_{ac,\boldsymbol{\theta},t}(\boldsymbol{\theta})|$ as

$$
\begin{aligned}
&\sup_{\boldsymbol{\theta}\in\Theta} |u_{b_q,t}(\boldsymbol{\theta})u_{t-k}(\boldsymbol{\theta}) + u_{b_q,t-k}(\boldsymbol{\theta})u_t(\boldsymbol{\theta}) - \tilde{u}_{b_q,t}(\boldsymbol{\theta})\tilde{u}_{t-k}(\boldsymbol{\theta}) - \tilde{u}_{b_q,t-k}(\boldsymbol{\theta})\tilde{u}_t(\boldsymbol{\theta})| \\
\leq & \sup_{\boldsymbol{\theta}\in\Theta_0} |u_{b_q,t} - \tilde{u}_{b_q,t}(\boldsymbol{\theta})| \sup_{\boldsymbol{\theta}\in\Theta_0} |u_{t-k}(\boldsymbol{\theta})| \\
& + \sup_{\boldsymbol{\theta}\in\Theta_0} |u_{t-k}(\boldsymbol{\theta}) - \tilde{u}_{t-k}(\boldsymbol{\theta})| \sup_{\boldsymbol{\theta}\in\Theta_0} |u_{b_q,t}(\boldsymbol{\theta}) - \tilde{u}_{b_q,t}(\boldsymbol{\theta})| \\
& + \sup_{\boldsymbol{\theta}\in\Theta_0} |u_{t-k}(\boldsymbol{\theta}) - \tilde{u}_{t-k}(\boldsymbol{\theta})| \sup_{\boldsymbol{\theta}\in\Theta_0} |u_{b_q,t}(\boldsymbol{\theta})|
\end{aligned}
$$

$$
\begin{aligned}
&+ \sup_{\boldsymbol{\theta}\in\Theta_0} |u_{b_q,t-k}(\boldsymbol{\theta}) - \tilde{u}_{b_q,t-k}(\boldsymbol{\theta})| \sup_{\boldsymbol{\theta}\in\Theta_0} |u_f t(\boldsymbol{\theta})| \\
&+ \sup_{\boldsymbol{\theta}\in\Theta_0} |u_{b_q,t-k}(\boldsymbol{\theta}) - \tilde{u}_{b_q,t-k}| \sup_{\boldsymbol{\theta}\in\Theta_0} |u_t(\boldsymbol{\theta}) - \tilde{u}_t(\boldsymbol{\theta})| \\
&+ \sup_{\boldsymbol{\theta}\in\Theta_0} |u_t(\boldsymbol{\theta}) - \tilde{u}_t(\boldsymbol{\theta})| \sup_{\boldsymbol{\theta}\in\Theta_0} |u_{b_q,t-k}(\boldsymbol{\theta})|,
\end{aligned}
$$

where the inequality follows from inequality (2.12). Using Minkowski and Cauchy-Schwartz inequalities and Lemmas C2 and C3, we see that this this is bounded in L_1-norm by

$$
\begin{aligned}
&C_1\rho^{T-t+1}(T-t+1) + C_2\rho^{T-t+k+1}\rho^{T-t+k+1}(T-t+k+1) \\
&+ C_3\rho^{T-t+1} + C_4\rho^{T-t+k+1}(T-t-k+1) \\
&+ C_5\rho^{T-t+k+1}(T-t+k+1)\rho^{T-t+1} + C_6\rho^{T-t+1},
\end{aligned}
$$

which clearly converges to zero as $T \to \infty$. Again, following the line of the proof of part (i), denote $g^{\bullet}_{\boldsymbol{\theta},t} = \sup_{\boldsymbol{\theta}\in\Theta_0} |g_{ac,\boldsymbol{\theta},T-t+1}(\boldsymbol{\theta}) - \tilde{g}_{ac,\boldsymbol{\theta},T-t+1}(\boldsymbol{\theta})|$, and note that

$$
\sum_{t=1}^{T} g^{\bullet}_{\boldsymbol{\theta},t} = \sum_{t=1}^{T} \sup_{\boldsymbol{\theta}\in\Theta_0} |g_{ac,\boldsymbol{\theta},t}(\boldsymbol{\theta}) - \tilde{g}_{ac,\boldsymbol{\theta},t}(\boldsymbol{\theta})|.
$$

Lemma A.2. in Meitz and Saikkonen (2011) implies that these sums have a.s. finite limits as $T \to \infty$ because

$$
\begin{aligned}
\|\gamma^t g^{\bullet}_{\boldsymbol{\theta},t}\| \leq &\gamma^t [C_1\rho^t t + C_2\rho^{2(t+k)}(t+k) + C_3\rho^t \\
&+ C_4\rho^{t+k}(t+k) + C_5\rho^{2(t+k)(t+k)} + C_6\rho^t],
\end{aligned}
$$

where the majorant converges to zero as $t \to \infty$ for some constant γ s.t. $\gamma > 1$ and $|\gamma\rho| < 1$. Thus

$$
\lim_{T\to\infty} \sum_{t=1}^{T} \sup |g_{ac,\boldsymbol{\theta},t}(\boldsymbol{\theta}) - \tilde{g}_{ac,\boldsymbol{\theta},t}(\boldsymbol{\theta})| < \infty \quad a.s.,
$$

and the majorant in inequality (2.15) has an a.s. limit zero as $T \to \infty$ and the result has been shown for the element corresponding to the derivative w.r.t. b_q. The same arguments are applicable to the rest of the elements as well.

In part (iv) we have to show that $\sum_{t=1}^{T} \sup_{\boldsymbol{\theta}\in\Theta_0} |g_{hs,\boldsymbol{\theta},t}(\boldsymbol{\theta}) - \tilde{g}_{hs,\boldsymbol{\theta},t}(\boldsymbol{\theta})|$ converges, and in order to do so, we show that the elements of the matrix $\sup_{\boldsymbol{\theta}\in\Theta_0} |g_{hs,\boldsymbol{\theta},t}(\boldsymbol{\theta}) - \tilde{g}_{hs,\boldsymbol{\theta},t}|$ converge to zero in L_1-norm. For the sake of brevity, we only consider the derivative w.r.t. parameters b_q and σ. The k^{th} element corresponding to the derivative w.r.t. b_q has a sup-norm

$$
\begin{aligned}
&\Bigg\| \sup_{\boldsymbol{\theta}\in\Theta_0} |2u_t(\boldsymbol{\theta})u_{b_q,t}(\boldsymbol{\theta})\kappa_{t-k}(\boldsymbol{\theta}) - 2\tilde{u}_t(\boldsymbol{\theta})\tilde{u}_{b_q,t}(\boldsymbol{\theta})\tilde{\kappa}_{t-k}(\boldsymbol{\theta}) \\
&-2u_{t-k}(\boldsymbol{\theta})u_{b_q,t-k}(\boldsymbol{\theta})\kappa_t(\boldsymbol{\theta}) - 2\tilde{u}_{t-k}(\boldsymbol{\theta})\tilde{u}_{b_q,t-k}(\boldsymbol{\theta})\tilde{\kappa}_t(\boldsymbol{\theta})| \Bigg\| \\
\le& \Bigg\| \sup_{\boldsymbol{\theta}\in\Theta_0} |u_t(\boldsymbol{\theta}) - \tilde{u}_t(\boldsymbol{\theta})| \Bigg\|_4 \Bigg\| \sup_{\boldsymbol{\theta}\in\Theta_0} |u_{b_q,t}(\boldsymbol{\theta})| \Bigg\|_4 \Bigg\| \sup_{\boldsymbol{\theta}\in\Theta_0} |\kappa_{t-k}(\boldsymbol{\theta})| \Bigg\|_2 \\
&+ \Bigg\| \sup_{\boldsymbol{\theta}\in\Theta_0} |u_{b_q,t}(\boldsymbol{\theta}) - \tilde{u}_{b_q,t}(\boldsymbol{\theta})| \Bigg\|_4 \Bigg\| \sup_{\boldsymbol{\theta}\in\Theta_0} |u_t(\boldsymbol{\theta}) - \tilde{u}_t(\boldsymbol{\theta})| \Bigg\|_4 \Bigg\| \sup_{\boldsymbol{\theta}\in\Theta_0} |\kappa_{t-k}(\boldsymbol{\theta})| \Bigg\|_2 \\
&+ \Bigg\| \sup_{\boldsymbol{\theta}\in\Theta_0} |u_{b_q,t}(\boldsymbol{\theta}) - \tilde{u}_{b_q,t}| \Bigg\|_4 \Bigg\| \sup_{\boldsymbol{\theta}\in\Theta_0} |u_t(\boldsymbol{\theta})| \Bigg\|_4 \Bigg\| \sup_{\boldsymbol{\theta}\in\Theta_0} |\kappa_{t-k}| \Bigg\|_2 \\
&+ \Bigg\| \sup_{\boldsymbol{\theta}\in\Theta_0} |u_t(\boldsymbol{\theta}) - \tilde{u}_t(\boldsymbol{\theta})| \Bigg\|_4 \Bigg\| \sup_{\boldsymbol{\theta}\in\Theta_0} |u_{b_q,t}(\boldsymbol{\theta})| \Bigg\|_4 \Bigg\| \sup_{\boldsymbol{\theta}\in\Theta_0} |\kappa_{t-k}(\boldsymbol{\theta}) - \tilde{\kappa}_{t-k}(\boldsymbol{\theta})| \Bigg\|_2 \\
&+ \Bigg\| \sup_{\boldsymbol{\theta}\in\Theta_0} |u_t(\boldsymbol{\theta}) - \tilde{u}_t(\boldsymbol{\theta})| \Bigg\|_4 \Bigg\| \sup_{\boldsymbol{\theta}\in\Theta_0} |u_{b_q,t}(\boldsymbol{\theta}) - \tilde{u}_{b_q,t}(\boldsymbol{\theta})| \Bigg\|_4 \Bigg\| \sup_{\boldsymbol{\theta}\in\Theta_0} |\kappa_{t_k}(\boldsymbol{\theta}) - \tilde{\kappa}_{t-k}(\boldsymbol{\theta})| \Bigg\|_2 \\
&+ \Bigg\| \sup_{\boldsymbol{\theta}\in\Theta_0} |\kappa_{t_k}(\boldsymbol{\theta}) - \tilde{\kappa}_{t_k}| \Bigg\|_2 \Bigg\| \sup_{\boldsymbol{\theta}\in\Theta_0} |u_t(\boldsymbol{\theta})| \Bigg\|_4 \Bigg\| \sup_{\boldsymbol{\theta}\in\Theta_0} |u_{b_q,t}(\boldsymbol{\theta})| \Bigg\|_4 \\
&+ \Bigg\| \sup_{\boldsymbol{\theta}\in\Theta_0} |u_{t-k}(\boldsymbol{\theta}) - \tilde{u}_{t-k}(\boldsymbol{\theta})| \Bigg\|_4 \Bigg\| \sup_{\boldsymbol{\theta}\in\Theta_0} |u_{b_q,t-k}(\boldsymbol{\theta})| \Bigg\|_4 \Bigg\| \sup_{\boldsymbol{\theta}\in\Theta_0} |\kappa_t| \Bigg\|_2 \\
&+ \Bigg\| \sup_{\boldsymbol{\theta}\in\Theta_0} |u_{t-k}(\boldsymbol{\theta}) - \tilde{u}_{t-k}(\boldsymbol{\theta})| \Bigg\|_4 \Bigg\| \sup_{\boldsymbol{\theta}\in\Theta_0} |u_{b_q,t-k}(\boldsymbol{\theta}) - \tilde{u}_{b_q,t-k}(\boldsymbol{\theta})| \Bigg\|_4 \Bigg\| \sup_{\boldsymbol{\theta}\in\Theta_0} |\kappa_t(\boldsymbol{\theta})| \Bigg\|_2 \\
&+ \Bigg\| \sup_{\boldsymbol{\theta}\in\Theta_0} |u_{b_q,t-k}(\boldsymbol{\theta}) - \tilde{u}_{b_q,t-k}(\boldsymbol{\theta})| \Bigg\|_4 \Bigg\| \sup_{\boldsymbol{\theta}\in\Theta_0} |u_{t-k}(\boldsymbol{\theta})| \Bigg\|_4 \Bigg\| \sup_{\boldsymbol{\theta}\in\Theta_0} |\kappa_t(\boldsymbol{\theta})| \Bigg\|_2
\end{aligned}
$$

$$
\begin{aligned}
&+\left\|\sup_{\boldsymbol{\theta}\in\Theta_0}|u_{t-k}(\boldsymbol{\theta})-\tilde{u}_{t-k}(\boldsymbol{\theta})|\right\|_4\left\|\sup_{\boldsymbol{\theta}\in\Theta_0}|u_{b_q,t-k}(\boldsymbol{\theta})|\right\|_4\left\|\sup_{\boldsymbol{\theta}\in\Theta_0}|\kappa_t(\boldsymbol{\theta})-\tilde{\kappa}_t(\boldsymbol{\theta})|\right\|_2\\
&+\left\|\sup_{\boldsymbol{\theta}\in\Theta_0}|\kappa_t(\boldsymbol{\theta})-\tilde{\kappa}_t(\boldsymbol{\theta})|\right\|_2\left\|\sup_{\boldsymbol{\theta}\in\Theta_0}|u_{t-k}(\boldsymbol{\theta})|\right\|_4\left\|\sup_{\boldsymbol{\theta}\in\Theta_0}|u_{b_q,t}(\boldsymbol{\theta})|\right\|_4,
\end{aligned}
$$

where the inequality follows from Cauchy-Schwartz and Minkowski inequalities and using inequality (2.12) twice. Using Lemmas C2 and C3 and the inequality (2.14), the majorant can be bounded by

$$
\begin{aligned}
&C_1(T-t+1)\rho^{T-t+1}+C_2(T-t+1)\rho^{2(T-t+1)}\\
&+C_3(T-t+1)\rho^{T-t+1}C_4\rho^{T-t+1}\\
&+C_5(T-t+1)\rho^{2(T-t+1)}+C_6(T-t+1)\rho^{T-t+1}\\
&+C_7(T-t+1)+C_8\rho^{2(T-t+1)}\\
&+C_9\rho^{T-t+1}+C_{10}\rho^{T-t+1}\\
&+C_{11}\rho^{T-t+k+1}+C_{12}(T-t+k+1)\rho^{2(T-t+k+1)}\\
&+C_{13}(T-t+k+1)\rho^{T-t+k+1}+C_{14}\rho^{T-t+k+1}\\
&+C_{15}\rho^{T-t+1},
\end{aligned}
$$

which converges to zero as $T\to\infty$.

In order to show the same result for the elements of the derivative matrix w.r.t. σ, note that the sup-norm of the k^{th} element is

$$
\begin{aligned}
&\left\|\sup_{\boldsymbol{\theta}\in\Theta_0}|-2\sigma(\kappa_t(\boldsymbol{\theta})-\tilde{\kappa}_t(\boldsymbol{\theta}))-2\sigma(\kappa_{t-k}(\boldsymbol{\theta})-\tilde{\kappa}_{t-k}(\boldsymbol{\theta}))|\right\|\\
\leq&\left\|\sup_{\boldsymbol{\theta}\in\Theta_0}|\sigma|\right\|_2\left(\left\|\sup_{\boldsymbol{\theta}\in\Theta_0}|\kappa_t(\boldsymbol{\theta})-\tilde{\kappa}_t(\boldsymbol{\theta})|\right\|_2+\left\|\sup_{\boldsymbol{\theta}\in\Theta_0}|\kappa_{t-k}(\boldsymbol{\theta})-\tilde{\kappa}_{t-k}(\boldsymbol{\theta})|\right\|_2\right)\\
\leq&C_1\rho^{T-t+1}+C_2\rho^{T-t+k+1},
\end{aligned}
$$

which converges to zero as $T\to\infty$. Applying these arguments to all elements in the derivative matrix and the techniques introduced in the previous parts, the convergence result can be shown elementwise. This completes the proof.

3 Maximum likelihood estimation for noninvertible ARMA model with α-stable errors[1]

3.1 Introduction

Financial time series data is often found to exhibit occasional large jumps around its average values. These large sudden movements may be asymmetric, meaning that large sudden crashes, for example, may be more common than sudden booms. These movements may seem like "spikes", or perhaps outliers, when financial time series data is plotted. Common methods in time series analysis often rely on the assumption of finite second moments of the observations. By doing so, in many cases estimators of the model parameters are shown to be $n^{1/2}$-consistent and asymptotically Gaussian; a results that follows using standard arguments applying a suitable central limit theorem. Then again, the aforementioned sudden jumps should not exist in light of this assumption on finite second moments.

[1]I am grateful for the guidance and suggestions of Markku Lanne, Mika Meitz, and Pentti Saikkonen. I am also thankful to Henri Nyberg and to all the participants at CFE-CMStatistics conference 2017, and HECER Econometrics Workshop, for useful comments. Detailed comments from Beth Andrews and Alain Hecq are gratefully acknowledged. I want to thank John Nolan for providing STABLE library for Matlab (http://academic2.american.edu/ jpnolan/stable/stable.html). Financial support from the Academy of Finland, Yrjö Jahnsson Foundation, and OP-Pohjola Research Foundation is gratefully acknowledged.

Jumps are not the only feature of financial data which is cumbersome for conventional linear time series models. Nonlinear behavior of the data is both well documented in the empirical financial literature, and also backed up by dynamic asset pricing theory (e.g. Singleton, 2009, Chapter 9). Clustering of the volatility of observations is perhaps the most common manifestation of these nonlinearities.

In this paper, we consider a simple univariate time series model that is potentially capable of controlling these feature in financial data: the noninvertible ARMA model with α-stable error term process. Noninvertible time series models have been shown to fit well in many applications to financial time series data, see for example, Breidt, Davis, and Trindade (2001), Andrews, Davis, and Breidt (2006), Lanne, Meitz, and Saikkonen (2013), and Huang and Pawitan (2000), and also Nyholm (2017). For example, these models are shown to control mild heteroskedasticity and they are capable of producing data that is only mildly autocorrelated, but still dependent in a nonlinear manner.

Forward looking nature of the noninvertible ARMA model relates closely to the (mixed causal) noncausal AR models. Both of these models can be written in a linear form of current, past, and future error terms. Gouriéroux and Zakoïan (2017) and Fries and Zakoïan (2019) shows how this forward looking presentation of the mixed causal-noncausal process is useful in modeling local explosive bubble patterns in commodity prices. They provide ordinary least square estimation theory for the AR parameters, as well as residual based diagnostic checks for the model selection.

Identification of this class of models is not possible under the assumption of a Gaussian error process, but α-stable distribution allows us to distinguish invertible and noninvertible models apart. For more information on the identification of the model, see Rosenblatt (2012), Chapter 1.

Maximum likelihood (ML) estimation in causal and invertible ARMA models with infinite variance has been considered by Davis, Knight, and Liu (1992), and Davis (1996), among others. Corresponding results for noncausal AR models were derived by Calder (1998) and Andrews, Calder, and Davis (2009). To the best of our knowledge, this paper is the first to consider ML estimation in noninvertible ARMA models with infinite variance errors. Related results for least absolute deviation estimation in general (possibly noninvertible) ARMA models were provided by Wu and Davis (2010), and for M-estimation by Wu (2013).

Our paper intersects with the papers by Calder (1998), and Andrews et al. (2009). They consider noncausal AR processes, whereas we consider causal and noninvertible ARMA models. Nevertheless, the estimation theory is very similar in both cases, and the asymptotics of the MA parameters are similar to those of the AR parameters in both cases. Large parts of the derivations in this paper are adopted from their work.

This paper overlaps substantially with Wu (2013), where M-estimation of a general ARMA model is considered. Our model is nested by their more general model specification, and ML estimation can be seen as a special case of their M-estimation procedure, although they do not consider ML estimation explicitly. There are, however, a few benefits in our approach. Our model specification is slightly different than the one in Wu (2013). In our specification, the observations are explicitly written in terms of future and current error terms, and the past observations, and we assume that AR polynomial satisfies the normal causality condition, and we write the MA polynomial in terms of the lead operator (in contrast to the standard way of writing it in terms of the lag operator) and assume that the roots of this polynomial situate outside the unit circle. In Chapter 3.2 we argue, that our specification simplifies the estimation and it has some practical benefits when hypothesis testing is considered. We will clarify these points when the model is introduced in detail in Section 3.2. Another difference is that we estimate the parameters of the α-stable distribution at the same time as the AR and MA parameters. ML estimation is a special case of their M-estimation, if we assumed that these error distribution parameters were known. Estimation of the distribution parameters comes in handy, since the asymptotic distribution of the AR and MA parameters is dependent on the tail parameter α of the stable distribution. To be precise, the tail parameter defines the rate of convergence of the estimators to their limiting distribution. Using the now proposed ML estimation method, we get the estimate for these parameters at once, and it can be used, for example, to calculate the confidence regions for the parameter estimates.

The estimation theory of ARMA models without finite second moments is based on the theory of weak convergence of point processes. Davis and Resnick (1985, 1986) provided weak convergence results for infinite order moving averages of stable distributed random variables. These results serves as a base for the estimation theory in ARMA literature, since these models have the MA(∞) representation. These results were used in estimation theory for ARMA models for example in ML estimation method of ARMA models without finite second moments are based on the theory of weak convergence of the point processes (e.g. Davis et al., 1992; Andrews et al., 2009). Using these methods, we are able to show the convergence in distribution of the estimators of the AR and MA parameters, as well as the estimators of the parameters of the stable distribution of the error process. The latter is $n^{1/2}$-

consistent and has an asymptotic Gaussian distribution, whereas the former converges at a faster rate to some non-standard distribution.

The remainder of this paper is organized as follows. Chapter 3.2 describes the noninvertible ARMA model and discusses the model specification by Meitz and Saikkonen (2013). It also recaps some of the key features of the α-stable distribution. In section 3 we give the main convergence results for the parameter estimators and provide an outline of the proof for the main theorem, with the most essential intermediate results. In Chapter 3.4 we illustrate the asymptotic properties in a small scale Monte Carlo simulation study and also apply the theory to trading volume data of Wal-Mart stock on New York Stock Exchange. SChapter 3.5 concludes. Lengthy steps for the proof of the main result are given in Appendix, as well as some assumptions and further notation. This paper comes with a supplementary appendix, where the detailed proofs of the main theorem can be found.

3.2 The noninvertible ARMA model with α-stable errors

3.2.1 The noninvertible ARMA model

The noninvertible process under consideration is

$$a_0(B)y_t = b_0(B^{-1})\varepsilon_t \tag{3.1}$$

with an AR polynomial $a_0(z) = 1 - a_{0,1}z - \cdots - a_{0,P}z^P$ and MA polynomial $b_0(z^{-1}) = 1 - b_{0,1}z^{-1} - \cdots - b_{0,Q}z^{-Q}$. B is a backshift operator (e.g. $B^k = y_{t-k}$ for $k = 0, \pm 1, \dots$). We assume that the following conditions for the roots of the polynomials $a_0(z)$ and $b_0(z^{-1})$ are satisfied:

$$a_0(z) \neq 0 \quad \text{for all } |z| \leq 1 \quad \text{and} \quad b_0(z^{-1}) \neq 0 \quad \text{for all } |z^{-1}| \leq 1. \tag{3.2}$$

Under these root conditions, the MA(∞) and AR(∞) representations of y_t and ε_t reads as

$$y_t = \frac{b_0(B^{-1})}{a_0(B)} = \sum_{j=-Q}^{\infty} \pi_{0,j}\varepsilon_{t-j} \quad \text{and} \quad \varepsilon_t = \frac{a_0(B)}{b_0(B^{-1})}y_t = \sum_{j=-P}^{\infty} \psi_{0,j}y_{t+j},$$

where the geometrically decaying sequences $\pi_{0,j}$ and $\psi_{0,j}$ are the coefficients of the Laurent series expansions of the rational polynomials $b_0(z^{-1})a_0(z)^{-1}$ and $a_0(z)b_0(z^{-1})^{-1}$ respectively.

The formulation of noninvertible ARMA(P,Q) model is adopted from Meitz and Saikkonen (2013). In this formulation, observation y_t is linearly dependent on its P lagged values, the current error term ε_t, and Q future error terms. This formulation differs slightly from presentations in the other literature (for example Lii and Rosenblatt, 1996). In our formulation, the dependence of y_t on the future error terms is made explicit by writing the MA polynomial in terms of the lead operator B^{-1}, whereas in the previous papers, noninvertibility is implied by using MA polynomial in terms of B, and instead of imposing the root condition (3.2), the polynomial is assumed to have its roots inside the unit circle. Although both formulations imply the same set of noninvertible ARMA models, the present formulation is preferable for two reasons. First, this formulation slightly simplifies the ML estimation, because the log-likelihood function (presented in the next section) is well defined also if any of the parameters gets the value zero. In our formulation, the log-likelihood function is absent of a term like $\ln|b_{0,Q}|$, whereas this terms occurs in the alternative presentation, see for example Lii and Rosenblatt (1996); Andrews et al. (2006). This also relates to the second reason: Testing for the true order of the MA polynomial would imply $b_{0,Q} = 0$, which would make the log-likelihood function undefined. These tests, however, are not considered in this paper, and are left for the future research.

We assume that the error terms of the model follow an α-stable distribution with infinite second moments. Let the iid sequence $\{\varepsilon_t\}$ have an α-stable distribution with $\alpha \in (0,2)$. This means, by definition, that there are sequences $\{\alpha_n\}$ and $\{\beta_n\}$ $(n = 1,2,\dots)$ s.t. $\alpha_n(\varepsilon_1 + \cdots + \varepsilon_n) + \beta_n \overset{d}{=} \varepsilon_1$.[2] Property 1.2.3 in Samoradnitsky and Taqqu (1994) generalizes this property for the infinite sums, and it follows that the process $y_t = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j}$ has also a stable distribution with the characteristic exponent α_0. Thus, the ARMA process y_t in (3.1) has finite moments up to α_0, $\mathrm{E}|y_t|^\delta < \infty$ for $\delta \in [0,\alpha_0)$. The tail probabilities of the α-stable distributions are characterized, for $\alpha_0 < 2$, as (see Property 1.2.15. in Samoradnitsky and Taqqu, 1994)

$$\lim_{x\to\infty} x^{\alpha_0} P(|\varepsilon_t| > x) = \tilde{c}(\alpha_0)\sigma_0^2, \quad \text{with} \tilde{c}(\alpha_0) = \left(\int_0^\infty t^{-\alpha_0}\sin(t)dt\right).$$

[2] Abbreviation $X \overset{d}{=} Y$ means that X and Y share the same distribution function.

The tails of the distribution behaves as $x^{-\alpha_0}$.

Other parameters that index the stable distributions are the location and scale parameters $\mu_0 \in \mathbb{R}$ and $|\sigma_0| \in (0, \infty)$. Parameter of symmetry is denoted by $|\beta_0| \leq 1$ and for $\beta_0 = 0$ the distribution is symmetric around μ_0. These parameters fully index the the characteristic function of the distribution and for now on we denote the index by $\lambda_0 = (\alpha_0, \beta_0, \sigma_0, \mu_0)$.

Although, in general, there is not a closed form presentation for the distribution function, there is a known characteristic function that can be used in the estimation (Nolan, 1997).

Class of α-stable distributions is large, so that this assumption is not very restrictive. For example, Gaussian distribution belongs to this class and $N(\mu_0, 2\sigma_0^2)$ is obtained by setting $\alpha_0 = 2$ and $\beta_0 = 0$. Since we are interested in processes with heavier tails, we should expect a distribution other than Gaussian. Gaussianity must also be assumed away in order to achieve the identification of the model. We get a symmetric Cauchy distribution by setting $\alpha_0 = 1$ and $\beta_0 = 0$. For these two distributions, there is a known closed form for the density functions.

This is the main technical reason for choosing to work with stable distributions. The scaled sum preserves the distribution of the summands, and we utilize this property when we derive asymptotic results for the estimators. Stable distributions are limits to scaled sums of iid random variables with heavy tails, much like Gaussian distribution serves as a limit of a scaled sums of iid random variables with finite second moments. It was pointed out by Calder (1998), that if the error term ε_t in our noninvertible ARMA model results additively from multiple sources, this distributional may provide a close approximation for this distribution.

3.3 Maximum likelihood estimation

3.3.1 Preliminaries

Our task is to estimate the parameter $\tau_0 = (\theta_0, \lambda_0)$ where $\theta_0 = (a_0, b_0)$ with $a_0 = (a_{0,1}, \ldots, a_{0,P})$ and $b_0 = (b_{0,1}, \ldots, b_{0,Q})$, and $\lambda_0 = (\alpha_0, \beta_0, \sigma_0, \mu_0)$. Sub-index zero is used to denote the true but unknown values of the parameters of (3.1). For all values $\tau = (\theta, \lambda)$ with $\theta = (a_1, \ldots, a_P, b_1, \ldots, b_Q) = (a, b)$ and $\lambda = (\alpha, \beta, \sigma, \mu)$, we define a counterpart of the process (3.1) as

$$a(B)y_t = b(B^{-1})u_t(\theta),$$

where $a(z) = 1 - a_1 z - \cdots - a_P z^P$ and $b(z^{-1}) = 1 - b_1 z^{-1} - \cdots - b_Q z^{-Q}$. In Assumption (1) in Appendix we define a permissible parameter space Θ_0 which obtains all the values of τ_0, for which our asymptotic results apply. For all the AR and MA parameters in this set Θ_0, root conditions similar to those in (3.2) are satisfied. Process $u_t(\theta)$ is a counterpart of the error process ε_t, and in Θ_0 it has the corresponding AR(∞) representation

$$u_t(\theta) = \frac{a(B)}{b(B^{-1})} y_t = \sum_{j=-P}^{\infty} \psi_j y_{t+j} \quad \text{for all } \theta \in \Theta_\theta,$$

with geometrically decaying coefficients ψ_j. For $\theta = \theta_0$ we have $u_t(\theta_0) = \varepsilon_t$. For the derivatives of $u_t(\theta)$ we have the following expression,

$$(\theta - \theta_0)^T \frac{\partial}{\partial \theta} u_t(\theta_0) \overset{def}{=} \sum_{j=-\infty}^{\infty} c_j(\theta) \varepsilon_{t-j}, \tag{3.3}$$

and the coefficients $c_j(\theta)$ are defined in (3.11) in Appendix. It can be seen, that for a fixed θ, the sequence $c_j(\theta)$ is a geometrically decaying, so $(\theta - \theta_0)\frac{\partial}{\partial \theta} u_t(\theta)$ is also in the domain of attraction of stable distribution with characteristic exponent α_0.

Using the sequence $u_t(\theta)$, we can define the log-likelihood function as

$$\mathscr{L}_n(\theta; \lambda) = \ln f(u_t(\theta); \lambda). \tag{3.4}$$

The sequence $u_t(\theta)$ is not feasible in practice since we do not observe the infinite future of the process y_t. For this reason we must use an approximate log-likelihood function in estimation. However, expression (3.4) is useful in the derivations of the main result.

We base our estimation on $n + P$ observations $\{y_t\}_{t=1-P}^{n}$. Solving for $u_t(\theta)$ would require the infinite future of y_t, so we have to approximate $u_t(\theta)$ by its feasible counterpart $\tilde{u}_t(\theta)$. To this end, we initialize by $\tilde{u}_{n+1}(\theta) = \cdots = \tilde{u}_{n+Q}(\theta) = 0$ and solve for $u_t(\theta)$ top-down, for $t = n, \ldots, 1$

$$\tilde{u}_t(\theta) = y_t - a_1 y_{t-1} - \cdots - a_P y_{t-P} + b_1 \tilde{u}_{t+1}(\theta) + \cdots + b_Q \tilde{u}_{t+Q}(\theta),$$

and obtain an observable set $\{\tilde{u}_t(\theta)\}_{t=1}^{n}$, which can be used in ML estimation.

The approximate log-likelihood function to be maximized is[3]

$$\tilde{\mathscr{L}}_n(\theta;\lambda) = \sum_{t=1}^{n} \ln f(\tilde{u}_t(\theta);\lambda). \tag{3.5}$$

The main contribution of this paper is to provide asymptotic results for the maximizer of the log-likelihood function, $\tilde{\tau}_n = (\tilde{\theta}_n, \tilde{\lambda}_n) = \arg\max_{\tau\in\Theta_0} \tilde{\mathscr{L}}_n(\theta;\lambda)$.

3.3.2 Asymptotic results

Before we state the main asymptotic results for the maximizer $\tilde{\tau}_n$, we introduce another random functions $\mathcal{Q}(v)$,

$$\mathcal{Q}_n(v,w) \stackrel{def}{=} \mathscr{L}_n(\theta_0 + n^{-1/\alpha_0}v; \lambda_0 + n^{-1/2}w) - \mathscr{L}_n(\theta_0;\lambda_0), \quad \text{and} \tag{3.6}$$

$$\mathcal{Q}(v) \stackrel{def}{=} \sum_{k=1}^{\infty}\sum_{j\neq 0}[\ln f(\varepsilon_{k,j} + \tilde{c}(\alpha_0)^{1/\alpha_0}\sigma_0 c_j(v)\delta_k\Gamma_k^{-1/\alpha_0};\lambda_0) - \ln f(\varepsilon_{k,j};\lambda_0)], \tag{3.7}$$

where

1. $\varepsilon_{k,j}$ is an iid sequence with $\varepsilon_{k,j} \stackrel{d}{=} \varepsilon_1$,
2. δ_k is an iid with $P(\delta_k = 1) = (1+\beta_0)/2$ and $P(\delta_k = -1) = 1 - (1+\beta_0)/2$,
3. $\Gamma_k = E_1 + \cdots + E_k$ where E_k is an iid series of exponential r.v.'s with mean one,
4. $\varepsilon_{k,j}$, δ_k and E_k are mutually independent, and
5. $c_j(\theta)$ is defined in (3.11).

Maximizing $\mathscr{L}_n(\theta;\lambda)$ in (3.5) w.r.t. θ and λ is equivalent to maximizing $\mathcal{Q}_n(v,w)$ in (3.6) w.r.t. $v = n^{1/\alpha_0}(\theta-\theta_0)$ and $w = n^{1/2}(\lambda-\lambda_0)$. It turns out, that in the derivations, the form (3.6) is easier to work with, than the log-likelihood $\mathscr{L}_n(\theta;\lambda)$ in (3.5), which is the objective function in practice. Random function $\mathcal{Q}(v)$ in (3.7) is not dependent on n. This random function defines part of the limiting distribution of the modified log-likelihood $\mathcal{Q}_n(v,w)$ on $\mathbb{C}(\mathbb{R}^{P+Q+4})$, the space of all the continuous functions on $\mathbb{R}^{P+Q+4}$. The following Proposition ensures, that the random function $\mathcal{Q}(v)$ has a unique maximizer ξ almost surely.

[3] For a detailed derivation of this, see Meitz and Saikkonen (2013).

Proposition 1. *Random function $\mathcal{Q}(v)$ in* (3.7) *is almost surely finite for all $v \in \mathbb{R}^{P+Q}$, and has a unique maximum.*

Proof. See Theorem 3.1. in Andrews et al. (2009), and Remark 2 in Davis et al. (1992). □

The next theorem gives the asymptotic distribution for the ML estimator of τ_0. Assumptions for this results are given in Assumption 1 in Appendix, and they are standard in the related literature (Wu, 2013; Calder, 1998). The parameter vector θ_0 must be in the interior of a permissible parameter space that contains all the parameter values that satisfy the root conditions in (3.2). Moreover, for $\alpha_0 \in (1,2)$, we only consider symmetric distributions, $\beta_0 = 0$. This assumption is needed in order to show some of the necessary convergences in the proofs of the theorem.

Theorem 1. *Let y_t be generated by process* (3.1) *and ε_t an iid sequence of α-stable distributed random variables with $\lambda \in \Theta_\lambda$ (defined in Assumption 1 in Appendix). Then, there exists a sequence of local maximizers $\tilde{\tau}_n$ of $\tilde{\mathscr{L}}_n(\theta;\lambda)$ s.t. as $n \to \infty$,*

$$n^{1/\alpha_0}(\tilde{\theta}_n - \theta_0) \xrightarrow{d} \xi \quad \textit{and} \quad n^{1/2}(\tilde{\lambda}_n - \lambda_0) \xrightarrow{d} Y \sim N(0, \mathbf{I}^{-1}(\lambda_0))$$

where ξ is a unique maximizer of $\mathcal{Q}(\cdot)$, ξ and Y are independent and $\mathbf{I}^{-1}(\lambda_0) = -E\left[\frac{\partial^2}{\partial\lambda\partial\lambda^T}\ln f(\varepsilon_1;\lambda_0)\right]$.

The proof follows from Lemmas 1-6 in Appendix, where the proofs of the Lemmas can also be found. We illustrate briefly the key ideas in the following.

Outline of the proof

Our goal is to show functional convergence of the modified likelihood function $\mathcal{Q}_n(v,w)$ on $\mathbb{C}(\mathbb{R}^{P+Q+4})$ to some limiting random function

$$\mathcal{Q}_n(v,w) \xrightarrow{d} \mathcal{Q}(v) - \frac{1}{2}w^T\mathbf{I}(\lambda_0)w + w^T\mathbf{N} \tag{3.8}$$

with $\mathcal{Q}_n(v,w)$ and $\mathcal{Q}(v)$ given in (3.6) and (3.7), $\mathbf{N} \sim \mathrm{N}(0, \mathbf{I}(\lambda_0))$ and $\mathbf{N}$ independent of $\mathcal{Q}(v)$. We begin by showing the weak convergence in (3.8) on $\mathbb{R}^{P+Q+4}$. This has been done in Lemmas 1-4. Proofs rely heavily on the convergence results of the point processes, and the methods are adopted form Calder (1998) and Andrews et al. (2009). Convergence results for the point processes in time series context has been developed for example in Davis and Resnick (1985) and Davis and Resnick (1986).

Lemma 5 establishes the tightness of $\mathcal{Q}_n(v, w)$. Using Theorem 7.1. in Billingsley (1999), convergence (3.8) on $\mathbb{R}^{P+Q+4}$ and the tightness on $\mathcal{Q}_n(v, w)$ implies the weak convergence (3.8) on $\mathbb{C}(\mathbb{R}^{P+Q+4})$, which was our aim to show.

In Lemma 6 we show a uniform convergence in probability of the unfeasible (3.4) and feasible (3.5) version of the log-likelihoods.

Unimodality of $\ln f(\cdot; \lambda_0)$ and strict convexity around θ_0 ensures that there are unique maximizers of $\mathcal{Q}_n(v, w)$ and $\mathcal{Q}(v) - \frac{1}{2} w^T \mathbf{I}(\lambda_0) w + w^T \mathbf{N}$: $(\tilde{v}_n, \tilde{w}_n)$ and $(\xi, \mathbf{I}^{-1}(\lambda_0)\mathbf{N})$. Lemma 2.2. in Davis et al. (1992) states the weak convergence of these maximizers,

$$(\tilde{v}_n, \tilde{w}_n) \xrightarrow{d} (\xi, \mathbf{I}^{-1}(\lambda_0)\mathbf{N}) \quad \text{on } \mathbb{R}^{P+Q+4}.$$

The result in Theorem 1 follows now by the relation between the maximizers of $\mathcal{Q}_n(v, w)$ and $\tilde{\mathscr{L}}_n(\theta, \lambda)$. The asymptotic properties of the maximizer $(\tilde{v}_n, \tilde{w}_n)$ are the same as those of $(n^{1/\alpha_0}(\tilde{\theta}_n - \theta_0), n^{1/2}(\tilde{\lambda}_n - \lambda_0))$. $\square$

3.4 Numerical results

3.4.1 Simulation study

Theoretical results derived in the previous section are illustrated by a simulation study. We have simulated data using (3.1) with α-stable distributed error terms using six different combination of model parameters. For each combination, 1,000 data sets with 250 observations have been simulated and noninvertible ARMA(1,1) model has been estimated to each data. The estimation results are illustrated in Table 3.1. A moderate number of observations have been selected to present a realistic setting in empirical applications. 250 observations is commonly encountered for example in low frequency financial and macro economic data.

The first set of results (the first six rows in Table 3.1) represents the estimation under infinite mean process ε_t ($\alpha = 0.8 < 1$). We study this process under autocorrelated observations ($a_{0,1} = 0.2$ and $b_{0,1} = 0.8$, left part) and non-correlated observations ($a_{0,1} = b_{0,1} = 0.5$, right part). All of the parameters are estimated with adequate accuracy since the mean of the estimates deviates only slightly from the true parameter values. Also the standard deviation of

	Empirical mean	Empirical std. dev.	Asymptotic std. dev.		Empirical mean	Empirical std. dev	Asymptotic std. dev
$a_{0,1} = 0.2$	0.200	0.004		$a_{0,1} = 0.5$	0.500	0.003	
$b_{0,1} = 0.8$	0.787	0.098		$b_{0,1} = 0.5$	0.500	0.003	
$\alpha_0 = 0.8$	0.798	0.058	0.072		0.798	0.057	0.072
$\beta_0 = 0$	-0.003	0.108	0.094		-0.006	0.100	0.094
$\sigma_0 = 1$	1.029	0.178	0.109		0.991	0.107	0.109
$\mu = 0$	-0.007	0.086	0.077		0.003	0.080	0.077
$a_{0,1} = 0.2$	0.196	0.071		$a_{0,1} = 0.5$	0.454	0.236	
$b_{0,1} = 0.8$	0.804	0.052		$b_{0,1} = 0.5$	0.459	0.245	
$\alpha_0 = 1.8$	1.805	0.094	0.089		1.801	0.091	0.088
$\beta_0 = 0$	-0.014	0.506	0.360		0.000	0.495	0.360
$\sigma_0 = 1$	0.998	0.061	0.057		0.994	0.058	0.057
$\mu = 0$	0.002	0.282	0.110		0.005	0.267	0.110
$a_{0,1} = 0.2$	0.194	0.065		$a_{0,1} = 0.5$	0.45	0.249	
$b_{0,1} = 0.8$	0.802	0.046		$b_{0,1} = 0.5$	0.458	0.261	
$\alpha_0 = 0.8$	0.797	0.089	0.086		0.794	0.088	0.086
$\beta_0 = 0.5$	0.515	0.402	0.331		0.472	0.431	0.331
$\sigma_0 = 1$	0.997	0.062	0.056		0.991	0.059	0.056
$\mu = 0$	0.021	0.276	0.110		0.012	0.167	0.110

Table 3.1: Estimation results for the parameters of noninvertible ARMA(1,1) in (3.1) with α-stable error term process with $n = 250$ observations. Number of simulations is 1,000.

the estimate of λ_0 is close to its asymptotic values. For the non-correlated case, the AR and MA parameters are estimated slightly smaller than their correct values. This is likely due to the fact that the likelihood function has another local maximum close to the origin. A small part of the estimates are drawn toward it which makes the mean of the sample slightly smaller than expected. This happens despite of the fact that the noninvertible ARMA model is identified even in the non-correlated case.

The same pattern repeats for the other combinations of the parameters as well. The middle part illustrates the case of infinite variance but finite mean ($\alpha_0 = 1.8 < 2$) and the last part illustrates the case of skewed error process ($\beta_0 = 0.5$) with infinite mean ($\alpha_0 = 0.8$).

3.4.2 Application to financial data

We follow the example of Andrews et al. (2009), Wu and Davis (2010), and Cui, Fisher, and Wu (2014) and study the daily volume traded of Wal-Mart stock on the New York Stock Exchange from December 1st, 2003 to December 31st, 2004. This period spans over 271 transaction days, which is similar in size than the sample sizes in the simulation study in the previous subsection.

Although the data may not be generated by a process with finite second moments, sample autocorrelation functions may be useful in illustrating some of its properties. In Figure 3.1 we have plotted the sample autocorrelation functions of the observations and the squared observations. Both of these series seem to exhibit some autocorrelation, which indicates that there are dependencies, both linear and nonlinear, potentially controllable with noninvertible ARMA(1,1) model. The top row of figure 3.1 shows that there are some considerable "spikes" in the logarithmic trading volume data. These points together makes the noninvertible ARMA model a reasonable starting point for modeling.

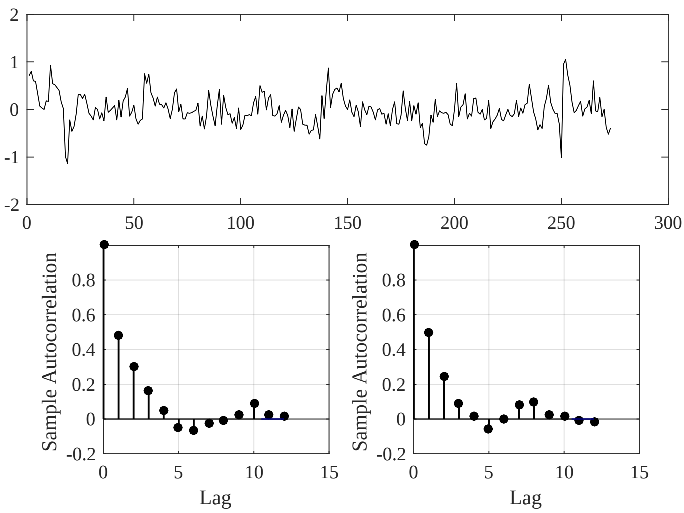

Figure 3.1: TOP: The demeaned logarithm of daily trade volume of Wal-Mart stock on the NYSE from December 1, 2003 to December 31, 2004. BOTTOM LEFT: Sample autocorrelation function for observed trading logarithmic trading volumes. BOTTOM RIGHT: Sample autocorrelation functions for squared observations of logarithmic trading volumes.

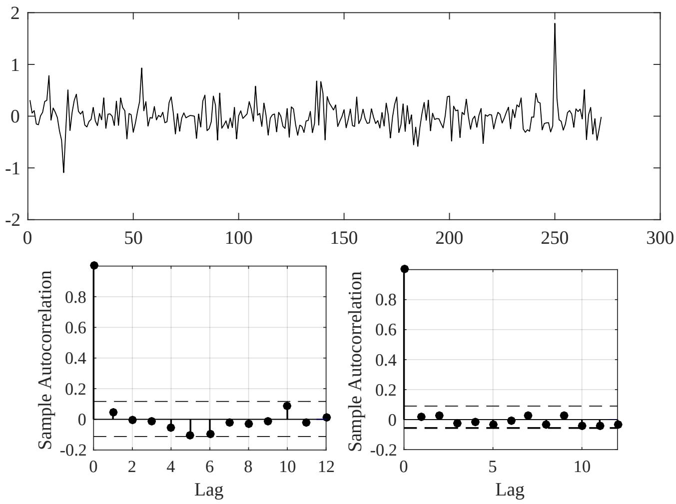

Figure 3.2: TOP: Residuals of the estimated model. BOTTOM LEFT: Sample autocorrelation functions for the residual series. BOTTOM RIGHT: Sample autocorrelation functions for the squared residuals.

The ML estimates of the noninvertible ARMA(1,1) model parameters $\theta_0 = (a_{0,1}, b_{0,1})$ and $\lambda_0 = (\alpha_0, \beta_0, \sigma_0, \mu_0)$ are

$$\tilde{\theta}_n = (0.720, 0.339) \quad \text{and} \quad \tilde{\lambda}_n = \Big(\underset{(.077)}{1.826}, \underset{(.335)}{0.601}, \underset{(.008)}{0.160}, \underset{(.017)}{-0.031} \Big).$$

Small numbers in the parenthesis indicate the standard errors of the estimated values. Estimated values enforces the findings of autocorrelated data as the AR and MA parameters are clearly different from zero and apart from each other. Estimated value of the tail index α_0 is smaller than two suggesting that the variance of the error process ε_t is infinite. This process is also skewed to the right as the estimate of β_0 is positive.

Residual series of the estimated model has been plotted in Figure 3.2 together with its sample autocorrelation functions and sample autocorrelation functions of its squared values. Confidence bounds to the sample autocorrelation pictures are obtained by simulating 10,000 iid series of $n = 275$ from the stable distribution with parameter values $\tilde{\lambda}_n$ and finding the 2.5 and 97.5 percentiles of the sample autocorrelation coefficients calculated from these series. The "spikiness" is visible in the residual series and the iid assumption on the residual series seems plausible by the sample autocorrelation functions.

According to these observations, the noninvertible ARMA(1,1) model with α-stable error distribution seems like an adequate model to capture the features of this trading volume data.

3.5 Conclusions

In this article we have derived asymptotic properties for the ML estimators of nonivertible ARMA model with α-stable errors with infinite variance. The rate of convergence of the estimators of AR and MA parameters is n^{1/α_0}, which is faster than the normal $n^{1/2}$ rate. The limiting distribution of the estimators is nonstandard. Parameters of the α-stable distributions converge to the Gaussian distribution at the conventional $n^{1/2}$ rate of convergence. Asymptotic results are derived by applying point process techniques.

Combination of noninvertible ARMA model and α-stable error process implies many appealing features. Noninvertible model allows us model some nonlinearities that are often encountered with financial time series data. For example, in our empirical example, we saw that mild heteroskedasticity in trading volume data of Wal-Mart stock, was successfully controlled by this model. At the same time, the model was capable of controlling for the autocorrelation in the data. Our results indicated, that the distribution of the trading volume data has heavy tails. If this is not incorporated in the estimation procedure, the asymptotics of the estimators might be distorted. Large visible jumps or spikes around the mean of the process can be seen as an evidence of the heavy tails. The α-stable distribution with infinite variance was also used to identify the noninvertible model from its invertible counterpart, which is not possible under assumption of Gaussian error term.

Hypothesis testing for the distribution parameters can be done in a usual manner, since the distribution is Gaussian and not dependent on the distribution of the AR and MA parameter estimators. Then again, hypothesis considering the AR and MA parameters is more difficult, because the limiting distribution is nonstandard. A bootstrap methods for estimating the distribution of the estimators of AR parameters in noncausal AR model is proposed in Andrews et al. (2009). These methods are outside the scope of this study, but they are reflected upon as a topic for future research.

References

Andrews, B., M. Calder, and R. A. Davis (2009). Maximum likelihood estimation for α-stable autoregressive processes. *Annals of Statistics 37*, 1946–1982.

Andrews, B., R. A. Davis, and F. J. Breidt (2006). Maximum likelihood estimation for all-pass time series models. *Journal of Multivariate Analysis 97*, 1638 – 1659.

Billingsley, P. (1999). *Convergence of probability measures*. Wiley, New York.

Breidt, F. J., R. A. Davis, and A. A. Trindade (2001). Least absolute deviation estimation for all-pass time series models. *Annals of statistics 29*, 919–946.

Calder, M. (1998). *Parameter estimation for noncausal and heavy-tailed autoregressive processes.* Ph. D. thesis, Departmen of Statistics. Colorado State University, Fort Collins, CO.

Cui, Y., T. J. Fisher, and R. Wu (2014). Diagnostic tests for non-causal time series with infinite variance. *Journal of Statistical Planning and Inference 147*, 117–131.

Davis, R. and S. Resnick (1985). Limit theory for moving averages of random variables with regularly varying tail probabilities. *Annals of Probability 13*, 179–195.

Davis, R. and S. Resnick (1986). Limit theory for the sample covariance and correlation functions of moving averages. *Annals of Statistics 14*, 533–558.

Davis, R. A. (1996). Gauss-newton and m-estimation for arma processes with infinite variance. *Stochastic Processes and Their Applications 63*, 75–95.

Davis, R. A., K. Knight, and J. Liu (1992). M-estimation for autoregressions with infinite variance. *Stochastic Processes and Their Applications 40*, 145–180.

Embrechts, P., C. Klüppelberg, and T. Mikosch (1997). *Modelling Extremal Events: for Insurance and Finance* (1 ed.). Applications of Mathematics 33. Springer-Verlag Berlin Heidelberg.

Feller, W. (1971). *An Introduction to Probability Theory and its Applications, Vol 2* (2 ed.). Wiley, New York.

Fries, S. and J.-M. Zakoïan (2019). Mixed causal-noncausal AR processes and the modelling of explosive bubbles. *Econometric Theory*. Forthcoming.

Gouriéroux, C. and J.-M. Zakoïan (2017). Local explosion modelling by non-causal process. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 79*, 737–756.

Huang, J. and Y. Pawitan (2000). Quasi-likelihood estimation of non-invertible moving average processes. *Scandinavian Journal of Statistics 27*, 689–702.

Kallenberg, O. (1983). *Random measures*. Akademie-Verlag.

Lanne, M., M. Meitz, and P. Saikkonen (2013). Testing for linear and nonlinear predictability of stock returns. *Journal of Financial Econometrics 11*, 682–705.

Lii, K.-S. and M. Rosenblatt (1996). Maximum likelihood estimation for non-gaussian nonminimum phase arma sequences. *Statistica Sinica 6*, 1–22.

Meitz, M. and P. Saikkonen (2013). Maximum likelihood estimation of a non-invertible arma model with autoregressive conditional heteroskedasticity. *Journal of Multivariate Analysis 114*, 227–255.

Nolan, J. P. (1997). Numerical calculation of stable densities and distribution functions. *Communications in Statistics. Stochastic Models 13*, 759–774.

Nyholm, J. (2017). Residual-based diagnostic tests for noninvertible ARMA models. Working Papers 416, Helsinki Center of Economic Research discussion papers.

Rosenblatt, M. (1985). *Stationary Sequences and Random Fields* (1 ed.). Birkhäuser Basel.

Rosenblatt, M. (2012). *Gaussian and non-Gaussian linear time series and random fields*. Springer Science & Business Media.

Samoradnitsky, G. and M. Taqqu (1994). *Stable non-Gaussian random processes: stochastic models with infinite variance*. Stochastic Modeling Series. Taylor & Francis.

Singleton, K. J. (2009). *Empirical dynamic asset pricing: model specification and econometric assessment*. Princeton University Press.

Wu, R. (2013). M-estimation for general arma processes with infinite variance. *Scandinavian Journal of Statistics 40*, 571–591.

Wu, R. and R. A. Davis (2010). Least absolute deviation estimation for general autoregressive moving average time-series models. *Journal of Time Series Analysis 31*, 98–112.

Appendix

3.A Assumptions and details for $u_t(\theta)$ and $\tilde{u}_t(\theta)$

Assumption 1. *The true AR and MA parameters belong to an interior of a permissible parameter space $\Theta_\tau = \Theta_a \times \Theta_b$ where*

$$\Theta_a = \Big\{(a_1, \ldots, a_P) \in \mathbb{R}^P \ ; \ a(z) \neq 0 \text{ for } |z| \leq 1\Big\}$$
$$\Theta_b = \Big\{(b_1, \ldots, b_Q) \in \mathbb{R}^Q \ ; \ b(z) \neq 0 \text{ for } |z^{-1}| \leq 1\Big\}, \quad \text{and}$$

Moreover, (a) for $\alpha_0 \leq 1$, then $\Theta_\lambda = \Theta_\alpha \times \Theta_\beta \times \Theta_\sigma \times \Theta_\mu$ with

$$\Theta_\alpha = \{\alpha_0 \in (0,1]\},\ \Theta_\beta = \{\beta_0 \in (-1,1)\},\ \Theta_\mu = \{\mu_0 \in \mathbb{R}\}, \text{ and}$$
$$\Theta_\sigma = \{\sigma_0 \in \mathbb{R}^+\}.$$

(b) If $\alpha_0 > 1$, then $\Theta_\lambda = \Theta_\alpha \times \Theta_\beta \times \Theta_\sigma \times \Theta_\mu$ with

$$\Theta_\alpha = \{\alpha_0 \in (1,2)\},\ \Theta_\beta = \{0\},\ \Theta_\mu = \mathbb{R}, \text{ and } \Theta_\sigma = \mathbb{R}^+.$$

The permissible parameter space is $\Theta = \Theta_\tau \times \Theta_\lambda$.

Process $u_t(\theta)$

The counterpart of $\varepsilon_t = u_t(\theta_0) = b_0(B^{-1})^{-1}a_0(B)y_t$ for $\theta \neq \theta_0$ is given by

$$u_t(\theta) = \frac{a(B)}{b(B^{-1})}y_t = \sum_{j=-P}^{\infty} \pi_j y_{t+j}$$

with $a(z) = a_1 z - \cdots - a_P z^P$ and $b(z^{-1}) = b_1 z^{-1} - \cdots - b_Q z^{-Q}$. For all $\theta \in \Theta_\theta$, where Θ_θ is defined in Assumption 1, $u_t(\theta)$ is well defined. Geometrically decaying sequence π_j is the coefficients of z_j of the Laurent series expansion of $a(z)b(z^{-1})^{-1} \overset{def.}{=} \pi(z)$.

Derivatives of $u_t(\theta_0)$ w.r.t. a_p, $(p = 1, \ldots, P)$ and b_q, $(q = 1, \ldots, Q)$ gives the elements of $u_{\theta,t}(\theta) \overset{def}{=} \frac{\partial}{\partial\theta} u_t(\theta)$,

$$u_{a_p,t}(\theta_0) \overset{def}{=} \tfrac{\partial}{\partial a_p} u_t(\theta_0) = -a_0(B)^{-1}\varepsilon_{t-p} = -\sum_{j=0}^{\infty} \psi_j^{(a)} \varepsilon_{t-p-j}, \text{ and} \tag{3.9}$$

$$u_{b_q,t}(\theta_0) \overset{def}{=} \tfrac{\partial}{\partial b_q} u_t(\theta_0) = b_0(B^{-1})^{-1}\varepsilon_{t+q} = \sum_{j=0}^{\infty} \psi_j^{(b)} \varepsilon_{t+j+q}. \tag{3.10}$$

Power series expansions of $a_0(z)^{-1} \overset{def}{=} \psi^{(a)}(z)$ and $b_0(z)^{-1} \overset{def}{=} \psi^{(b)}(z)$ defines $\psi_j^{(a)}$ and $\psi_j^{(b)}$ as the coefficients of z^j respectively. These power series are well defined, and it follows that the coefficients $\psi_j^{(a)}$ and $\psi_j^{(b)}$ are geometrically decaying as $j \to \infty$.

Utilizing the notation $\theta = (a_1, \ldots, a_P, b_1, \ldots, b_Q)$, we have

$$\begin{aligned}(\theta-\theta_0)^T \tfrac{\partial}{\partial\theta} u_t(\theta_0) =& (a_1 - a_{0,1}) \tfrac{\partial}{\partial a_1} u_t(\theta_0) + \cdots + (a_P - a_{0,P}) \tfrac{\partial}{\partial a_P} u_t(\theta_0) \\ &+ (b_1 - b_{0,1}) \tfrac{\partial}{\partial b_1} u_t(\theta_0) + \cdots + (b_Q - b_{0,Q}) \tfrac{\partial}{\partial b_Q} u_t(\theta_0).\end{aligned}$$

Let us use the sum expressions in (3.9) and (3.10) and collect the coefficients of all ε_{t-j}, $j = \ldots, -1, 0, 1, \ldots$ we get the expression

$$(\theta-\theta_0)^T \tfrac{\partial}{\partial\theta} u_t(\theta_0) \overset{def}{=} \sum_{j=-\infty}^{\infty} c_j(\theta-\theta_0)\varepsilon_{t-j}$$

with

$$c_j(\theta-\theta_0) = \begin{cases} -\sum_{p=1}^{P}(a_p - a_{0,p})\psi_{0,j-p}^{(a)} & \text{for } j \geq P \\ -\sum_{p=1}^{j}(a_p - a_{0,p})\psi_{0,j-p}^{(a)} & \text{for } 1 \leq j < P \\ 0 & \text{for } j = 0 \\ \sum_{q=1}^{j}(b_q - b_{0,q})\psi_{0,q-j}^{(b)} & \text{for } -Q < j \leq -1 \\ \sum_{q=1}^{Q}(b_q - b_{0,q})\psi_{0,q-j}^{(b)} & \text{for } j \leq -Q \end{cases}. \tag{3.11}$$

Because $\psi_{0,j}^{(a)}$ and $\psi_{0,j}^{(b)}$ are geometrically decaying sequences, it is not hard to show that also $c_j(\theta)$ decays to zero at the geometric rate as $j \to \infty$.

For the derivation of the second derivatives, note that $u_{a_p,t}(\theta) =$

$-b_0(B^{-1})^{-1}y_{t-p}$ for $p = 1, \ldots, P$, so $\frac{\partial^2}{\partial a_p \partial a_{p'}} u_t(\theta) = 0$ for $p, p' = 1, \ldots, P$. From $a(B)y_t = b(B^{-1})u_t(\theta)$ we obtain $0 = \frac{\partial}{\partial b_q} b(B^{-1})u_t(\theta) = -B^{-q}u_t(\theta) + b(B^{-1})u_{b_q,t}(\theta)$ for $q = 1, \ldots, Q$. Taking derivatives w.r.t. a_p, $p = 1, \ldots, P$ and $b_{q'}$, $q' = 1, \ldots, Q$,

$$\frac{\partial^2}{\partial a_p \partial b_q} u_t(\theta) = b(B^{-1})^{-1} u_{a_p,t+q}(\theta)$$
$$\frac{\partial^2}{\partial b_q \partial b'_q} u_t(\theta) = b(B^{-1})^{-1} u_{b_{q'},t+q}(\theta) - B^{-q'} b(B^{-1})^{-2} u_{t+q}(\theta).$$

Rearranging terms gives

$$\frac{\partial^2}{\partial a_p \partial b_q} u_t(\theta) = -b(B^{-1})^{-1} a(B)^{-1} u_{t+q+p}(\theta) \quad \text{and}$$
$$\frac{\partial^2}{\partial b_q \partial b_{q'}} u_t(\theta) = 2b(B^{-1})^{-2} u_{t+q+q'}(\theta).$$

Lemma A1. *For all $\eta \in (0, \alpha_0)$, and $\theta \in \Theta_\theta$ described in Assumption 1,*

$$(i)\ E[|u_t(\theta)|^\eta] < \infty, \quad (ii)\ E[|u_{a_p,t}(\theta)|^\eta] < \infty,$$
$$(iii)\ E[|u_{b_q,t}(\theta)|^\eta] < \infty, \quad (iv)\ E[|u_{a_p,a_{p'},t}(\theta)|^\eta] < \infty,$$
$$(v)\ E[|u_{a_p,b_q,t}(\theta)|^\eta] < \infty, \quad \text{and} \quad (vi)\ E[|u_{b_q,b_{q'},t}(\theta)|^\eta] < \infty.$$

Proof. Results follow by the definition of Θ_θ, the series presentations given above, and Lemmas A.1. and A.2. in Meitz and Saikkonen (2013). □

Feasible counterpart of $u_t(\theta)$

In practice, we do not observe the process $u_t(\theta)$, so they can not be used for maximizing the likelihood. For this reason, in practice, we use feasible counterpart $\tilde{u}_t(\theta)$ instead. Set $\tilde{u}_{n+1}(\theta) = \cdots = \tilde{u}_{n+Q}(\theta) = 0$ and, using observations $\{y_t\}_{t=1-P}^{n}$, solve recursively for $\tilde{u}_n(\theta), \ldots, \tilde{u}_1(\theta)$ as

$$\begin{aligned}
\tilde{u}_n(\theta) &= a(B)y_n \\
\tilde{u}_{n-1}(\theta) &= a(B)y_{n-1} + b_1\tilde{u}_n(\theta) \\
&\vdots \\
\tilde{u}_{n-Q}(\theta) &= a(B)y_{n-Q} + b_1\tilde{u}_{n-Q+1}(\theta) + \cdots + b_Q\tilde{u}_n(\theta)
\end{aligned}$$

$$\vdots$$
$$\tilde{u}_1(\theta) = a(b)y_1 + b_1\tilde{u}_2(\theta) + \cdots + b_Q\tilde{u}_{1+Q}(\theta),$$

or as in Andrews et al. (2009), for $t = 1, \ldots, n$,

$$\tilde{u}_t(\theta) = \sum_{j=0}^{n-t} \psi_j^{(b)} a(B) y_{t+j}.$$

The feasible quantities $\tilde{u}_t(\theta)$ gets closer to its theoretical counterpart $u_t(\theta) = \sum_{j=0}^{\infty} \psi_j^{(b)} a(B) y_{t+j}$, as $n \to \infty$ but note that the values at the both ends of the time line can differ substantially. The difference between these quantities is

$$u_t(\theta) - \tilde{u}_t(\theta) = \sum_{j=n-t+1}^{\infty} \psi^{(b)} a(B) y_{t+j}. \tag{3.12}$$

The derivatives of $\tilde{u}_t(\theta)$ are given by (see p. 251 Meitz and Saikkonen, 2013)

$$\tilde{u}_{a_p,t}(\theta) = -\sum_{j=0}^{n-t} \psi_j^{(b)} y_{t-p+j} \quad \text{and} \quad \tilde{u}_{b_q,t}(\theta) = \sum_{j=0}^{n-t} \psi_j^{(b)} \tilde{u}_{t+q+j}(\theta). \tag{3.13}$$

Lemma A2. *For all $\eta \in (0, \alpha_0)$, for any $m_n = O(n^{1/2})$, $v \in \mathbb{R}^{P+Q}$, and $\theta \in \Theta_\theta$ defined in Assumption 1,*

$$\begin{aligned}
&(i)\ E[|\tilde{u}_t(\theta)|^\eta] < \infty, \quad (ii)\ E[|\tilde{u}_{a_p,t}(\theta)|^\eta] < \infty,\\
&(iii)\ E[|\tilde{u}_{b_q,t}(\theta)|^\eta] < \infty, \quad (iv)\ E[|\tilde{u}_{a_p a_{p'},t}(\theta)|^\eta] < \infty\\
&(v)\ E[|\tilde{u}_{a_p b_q,t}(\theta)|^\eta] < \infty, \quad (vi)\ E[|\tilde{u}_{b_q b_{q'},t}(\theta)|^\eta] < \infty,\\
&(vii)\ E\left[\left(\sum_{t=m_n}^{n-m_n} |\tilde{u}_t(\theta_0) - u_t(\theta_0)|\right)^\eta\right] \to 0, \quad \textit{and}\\
&(viii)\ E\left[\left(\sum_{t=m_n}^{n-m_n} v^T |\tilde{u}_{\theta,t}(\theta_0) - u_{\theta,t}(\theta_0)|\right)^\eta\right] \to 0.
\end{aligned}$$

Proof. Proof is given in a supplementary appendix. □

3.B Details for the proof of Theorem 1

The first three Lemmas in this section are in order to establish the weak convergence in (3.8) on $\mathbb{R}^{P+Q+4}$. To this end, for any fixed $v \in \mathbb{R}^{P+Q}$ and $w \in \mathbb{R}^4$, we introduce a random function[4]

$$\begin{aligned}\mathcal{Q}_n^*(v,w) &\overset{def}{=} \sum_{t=1}^{n}[\ln\ f(\varepsilon_t + n^{-1/\alpha_0} v^T u_{\theta,t}(\theta_0);\lambda_0) - \ln\ f(\varepsilon_t;\lambda_0)] \\ &\quad + n^{-1/2} w^T \sum_{t=1}^{n} e_{\lambda,t}(\varepsilon_t;\lambda_0) \\ &\overset{def}{=} \mathcal{Q}_n^+(v) + N_n(w)\end{aligned} \tag{3.14}$$

and show that

Lemma B3. *For any $v \in \mathbb{R}^{P+Q}$ and $w \in \mathbb{R}^4$,*

$$\mathcal{Q}_n(v,w) - \mathcal{Q}_n^*(v,w) = \frac{1}{2} w^T \mathbf{I}(\lambda_0) w + o_p(1).$$

Proof. Proofs for the Lemmas in this section are given in supplementary appendix. □

The next step is to show that $\mathcal{Q}^*(v,w)$ converges weakly for any $(v,w) \in \mathbb{R}^{P+Q+4}$. In order to do so, we must make use of the weak convergence results of the point processes. [5] Let us introduce two new random functions,

$$\begin{aligned}\mathcal{W}_{m,n}(\cdot) &\overset{def}{=} \sum_{t=1}^{n} I_{\left(\varepsilon_t, c(\alpha_0)^{-1}\sigma_0^{-1} n^{-1/\alpha_0} \boldsymbol{\epsilon}_{t,m}\right)}(\cdot) \quad \text{and} \\ \mathcal{W}_m(\cdot) &\overset{def}{=} \sum_{k=1}^{\infty} \sum_{j\in\{-m,\ldots m\}\backslash 0} I_{\left(\varepsilon_{k,j}, \iota_j \delta_k \Gamma_k^{-1/\alpha_0}\right)}(\cdot),\end{aligned}$$

where $\boldsymbol{\epsilon}_{t,m} = (\varepsilon_{t-m}, \ldots, \varepsilon_{t-1}, \varepsilon_{t+1}, \ldots, \varepsilon_{t+m})$,

$$\iota_j = (\underbrace{0,\ldots,0}_{m+j}, 1, \underbrace{0,\ldots,0}_{m-j}),$$

[4] Remember the expression $u_t(\theta_0) = \varepsilon_t$.

[5] Our treatment of point processes is adapted from Calder (1998), Chapter 2.3. Textbook material can be found in Embrechts, Klüppelberg, and Mikosch (1997) and Kallenberg (1983).

a $(2m \times 1)$ vector where the 1 is located at the $(m+j+1)^{th}$ element for $j<0$, and $(m+j)^{th}$ element for $j>0$. δ_k is an iid random variable with $P(\delta=1)=(1+\beta_0)/2$ and $P(\delta_k=-1)=1-P(\delta_k=1)$, and $\Gamma_k=E_1+\cdots+E_k$ with E_i an iid exponential random variable for $i=1,\ldots,\infty$. These functions define a random measure for any (compact) set $E\in\mathbb{R}^2$, for which $\mathcal{W}_{m,n}(E)<\infty$ and $\mathcal{W}_m(E)<\infty$. Following Andrews et al. (2009), we start by providing a finite dimensional convergence result for point processes $\mathcal{W}_{m,n}(\cdot)$ for any fixed sets A that makes the mapping measurable. We also show that limiting distribution of $N_n(w)$ is independent from the point processes distribution.

Lemma B4. *For any fixed (relatively compact subset) $A_s\in\mathbb{R}\times(\mathbb{R}^{2m}\setminus\mathbf{0})$ of the form $A_s=(\underline{a}_{s,0},\overline{a}_{s,0}]\times(\underline{a}_{s,-m},\overline{a}_{s,-m}]\times\cdots\times(\underline{a}_{s,-1},\overline{a}_{s,-1}]\times(\underline{a}_{s,1},\overline{a}_{s,1}]\times\cdots\times(\underline{a}_{s,m},\overline{a}_{s,m}]$, and for any $w\in\mathbb{R}^4$,*

$$(\mathcal{W}_{m,n}(A_1),\ldots,\mathcal{W}_{m,n}(A_l),N_n(w))\xrightarrow{d}(\mathcal{W}_m(A_1),\ldots,\mathcal{W}_m(A_l),w^T\mathbf{N}),$$

as $n\to\infty$, and $\mathcal{W}_m(A_k)$ is independent of $w^T\mathbf{N}$ for all $k=1,\ldots,l$.

This Lemma is essentially the same as Lemma A.9. in Andrews et al. (2009), and we can see that the simultaneous estimation of the distributional parameters λ with the AR and MA parameters θ complicates the estimation theory slightly, because we have to take care of the joint distribution of the parameters θ and λ (see for example Wu (2013)).

To see how the previous weak convergence result comes in useful in our context, we define two random functions

$$\mathcal{W}_n^*(\cdot)\stackrel{def}{=}\sum_{t=1}^{n}I_{\left(\varepsilon_t,\ n^{-1/\alpha_0}\sum_{j=1}^{\infty}c_j(v)\varepsilon_{t-j},\ n^{-1/\alpha_0}\sum_{j=1}^{\infty}c_{-j}(v)\varepsilon_{t+j}\right)}(\cdot)\qquad\text{and}$$

$$\mathcal{W}^*(\cdot)\stackrel{def}{=}\sum_{t=1}^{\infty}\left[\sum_{j=-1}^{-\infty}I_{\left(\varepsilon_{k,j},\ c(\alpha_0)^{1/\alpha_0}\sigma_0c_j(v)\delta_k\Gamma_k^{-1/\alpha_0},\ 0\right)}(\cdot)\right.$$
$$\left.+\sum_{j=1}^{\infty}I_{\left(\varepsilon_{k,j},\ 0,\ c(\alpha_0)^{1/\alpha_0}\sigma_0c_j(v)\delta_k\Gamma_k^{-1/\alpha_0}\right)}(\cdot)\right].$$

If we define a continuous mapping

$$(\varepsilon_{t-m},\ldots,\varepsilon_t,\ldots,\varepsilon_{t+m})$$

$$\mapsto \left(\varepsilon_t,\ c(\alpha_0)^{1/\alpha_0} \sigma_0 \sum_{j=-1}^{-m} c_j(v) \varepsilon_{t-j},\ c(\alpha_0)^{1/\alpha_0} \sigma_0 \sum_{j=1}^{m} c_j(v) \varepsilon_{t-j} \right),$$

and apply this together with continuous mapping theorem to the random vector

$$(\mathcal{W}_{m,n}(A_1), \ldots, \mathcal{W}_{m,n}(A_l))$$

in Lemma B4, and let $m \to \infty$, we find that

$$(\mathcal{W}_n^*(A_1), \ldots, \mathcal{W}_n^*(A_l)) \xrightarrow{d} (\mathcal{W}^*(A_1), \ldots, \mathcal{W}^*(A_l)) \tag{3.15}$$

for all A_s, $s = 1, \ldots, l$ defined in Lemma B4.

Next we make use of definition 5.2.1. in Embrechts et al. (1997). The weak convergence of the finite dimensional random vector (3.15) for all $A_s \in \mathbb{R} \times (\mathbb{R}^{2m} \setminus \mathbf{0})$ implies the weak convergence of $\mathcal{W}_n^*(\cdot)$ to $\mathcal{W}^*(\cdot)$ on $\mathbb{R}$. This in turn implies that for all positive valued functions g on $E = \mathbb{R} \times (\mathbb{R}^{2m} \setminus \mathbf{0})$, (Remark 3 in Embrechts et al., 1997, p. 234))

$$\int_E g(\cdot) d\mathcal{W}_n^* \xrightarrow{d} \int_E g(\cdot) d\mathcal{W}^*,$$

and the integrals simplifies to

$$\begin{aligned} &\sum_{t=1}^{n} g\left(\varepsilon_t,\ n^{-1/\alpha_0} \sum_{j=1}^{\infty} c_j(v) \varepsilon_{t-j},\ n^{-1/\alpha_0} \sum_{j=1}^{\infty} c_{-j}(v) \varepsilon_{t+j} \right) \\ &\xrightarrow{d} \sum_{k=1}^{\infty} \sum_{j=1}^{\infty} g((\varepsilon_{k,j},\ c(\alpha_0)^{1/\alpha_0} \sigma_0 c_j(v) \delta_k \Gamma_k^{-1/\alpha_0},\ 0) \\ &\qquad\qquad + g(\varepsilon_{k,-j},\ 0,\ c(\alpha_0)^{1/\alpha_0} \sigma_0 c_{-j}(v) \delta_k \Gamma_k^{-1/\alpha_0})). \end{aligned} \tag{3.16}$$

We can now see how to make statements about the weak convergence of the likelihood function at fixed (v, w) on $\mathbb{R}^{P+Q+4}$. Recalling (3.7) and (3.14), it would be tempting to set $g(x, y, z) = \ln f(x + y + z\ ;\ \lambda_0) - \ln f(x\ ;\ \lambda_0)$ and find out that $(\mathcal{Q}_n^+(v), N_n(w)) \xrightarrow{d} (\mathcal{Q}(v), w^T \mathbf{N})$, from which we could postulate, using (3.16), that

$$\mathcal{Q}_n^*(v) = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} \mathcal{Q}_n^+(v) \\ N_n(w) \end{bmatrix} \xrightarrow{d} \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} \mathcal{Q}(v) \\ w^T \mathbf{N} \end{bmatrix}$$

$$\Leftrightarrow \mathcal{Q}_n(v,w) \xrightarrow{d} \mathcal{Q}(v) + \frac{1}{2}w^T\mathbf{N} - w^T\mathbf{I}(\lambda_0) \tag{3.17}$$

on $\mathbb{R}^{P+Q+4}$. The reason why this is not straight forward is that function $g(x)$ does not have a compact support on E. That is, there is no compact $K \subset E$ s.t. $g(x) = 0$ for all $x \in K^c$ (the complement of K). The way to get around this problem is to use

$$\begin{aligned}\tilde{g}(x,y,z) = &[\ln f(x+y+z\ ;\ \lambda_0) - \ln f(x\ ;\ \lambda_0)]\\ &\times I\{|x| \le M\}\, I\{(|y| > \epsilon) \cup (|z| > \epsilon)\}\end{aligned}$$

for some possibly large $M < \infty$ and small $\epsilon > 0$. To see why the convergence in (3.17) is attained using g instead of $\tilde{g}$, we refer to Theorem 3.2. in Billingsley (1999) and the following two lemmas.

Lemma B5. *For*

$$\begin{aligned}&\tilde{g}(x,y,z)\\ &= [\ln f(x+y+x\ ;\ \lambda_0) - \ln f(x\ ;\ \lambda_0)]\, I\{|x| \le M\}\, I\{(|y| > \epsilon) \cup (|z| > \epsilon)\}\end{aligned}$$

and for any fixed $v \in \mathbb{R}^{P+Q}$,

$$\begin{aligned}\sum_{k=1}^{\infty}\sum_{j=1}^{\infty} &\tilde{g}((\varepsilon_{k,j}, c(\alpha_0)^{1/\alpha_0}\sigma_0 c_j(v)\delta_k\Gamma_k^{-1/\alpha_0},\ 0)\\ &+ \tilde{g}(\varepsilon_{k,-j}, 0, c(\alpha_0)^{1/\alpha_0}\sigma_0 c_{-j}(v)\delta_k\Gamma_k^{-1/\alpha_0})) \xrightarrow{p} \mathcal{Q}(v)\end{aligned}$$

on $\mathbb{R}^{P+Q}$ *as* $M \to \infty$ *and* $\epsilon \to 0$,

and

Lemma B6. *For* $g(x,y,z) = \ln f(x+y+z;\lambda_0) - \ln f(x;\lambda_0)$, *for any fixed* $v \in \mathbb{R}^{P+Q}$, *and for any* $\kappa > 0$,

$$\lim_{M\to\infty}\lim_{\epsilon\to 0}\limsup_{n\to\infty}$$
$$P\left(\left|\mathcal{Q}_n^+(v) - \sum_{t=1}^{n}\tilde{g}\left(\varepsilon_t, n^{-1/\alpha_0}\sum_{j=1}^{\infty}c_j(v)\varepsilon_{t-j}, n^{-1/\alpha_0}\sum_{j=1}^{\infty}c_{-j}(v)\varepsilon_{t+j}\right)\right| > \kappa\right) = 0.$$

That is, let

$$X_{n,\epsilon,M} = \sum_{t=1}^{n} \tilde{g}\left(\varepsilon_t,\ n^{-1/\alpha_0}\sum_{j=1}^{\infty} c_j(v)\varepsilon_{t-j},\ n^{-1/\alpha_0}\sum_{j=1}^{\infty} c_{-j}(v)\varepsilon_{t+j}\right) \quad \text{and}$$

$$Z_{\epsilon,M} = \sum_{k=1}^{\infty}\sum_{j=1}^{\infty} \tilde{g}((\varepsilon_{k,j}, c(\alpha_0)^{1/\alpha_0}\sigma_0 c_j(v)\delta_k\Gamma_k^{-1/\alpha_0}, 0) + \tilde{g}(\varepsilon_{k,-j},\ 0,\ c(\alpha_0)^{1/\alpha_0}\sigma_0 c_{-j}(v)\delta_k\Gamma_k^{-1/\alpha_0})),$$

then $X_{n,\epsilon,M} \xrightarrow{d} Z_{\epsilon,M}$ on $\mathbb{R}^{P+Q}$ as $n \to \infty$ by the previous arguments. In Lemma B5 we showed that $Z_{\epsilon,M} \xrightarrow{d} \mathcal{Q}(v)$ on $\mathbb{R}^{P+Q}$ as $M \to \infty$ and $\epsilon \to 0$. By Lemma B6 we have that $\lim_{M\to\infty}\lim_{\epsilon\to 0}\limsup_{n\to\infty} P(|X_{n,\epsilon,M} - \mathcal{Q}_n^+(v)| > \kappa) = 0$ for all $\kappa > 0$. Theorem 3.2. in Billingsley (1999) then implies that $\mathcal{Q}_n^+(v) \xrightarrow{d} Q(v)$ on $\mathbb{R}^{P+Q}$.

Weak convergence on $\mathbb{C}(\mathbb{R}^{P+Q+4})$ follows by the finite dimensional convergence and tightness of $\mathcal{Q}_n(v,w)$ on $\mathbb{C}(\mathbb{R}^{P+Q+4})$, which is the same as tightness on $\mathbb{C}(\mathbb{K})$ for any $\mathbb{K} = [-K,K]^4$.

Lemma B7. *For any $M, K > 0$ and $\kappa > 0$,*

$$\lim_{\epsilon\to 0^+}\limsup_{n\to\infty} P\left(\sup_{\substack{||v'-v''||<\epsilon \\ ||w'-w''||<\epsilon \\ ||v'||,||v''||,||w'||,||w''||<K}} \left|\mathcal{Q}_n(v',w') - \mathcal{Q}_n(v'',w'')\right| > \kappa\right) = 0.$$

Tightness of $\mathcal{Q}_n(v,w)$ together with the finite dimensional convergence (3.17) implies the weak convergence

$$\mathcal{Q}_n(v,w) \xrightarrow{d} \mathcal{Q}(v) - \frac{1}{2}w^T\mathbf{I}(\lambda_0)w + w^T\mathbf{N}$$

on $\mathbb{C}(\mathbb{R}^{P+Q+4})$ by Theorem 7.1 in Billingsley (1999).

Finally, we have to show that the initialization in calculating $\tilde{u}_t(\theta)$ does not alter our asymptotic results. To this end, we show that

Lemma B8. *For any* $(v, w) \in \mathbb{R}^{P+Q+4}$*, as* $n \to \infty$*,*

$$\sup_{v,w} \left| \tilde{\mathcal{Q}}_n(v, w) - \mathcal{Q}_{n,m}(v, w) \right| = o_p(1),$$

where $\tilde{\mathcal{Q}}_{m,n}(v, w) = \tilde{\mathscr{L}}_n(\theta_0 + n^{-1/\alpha_0} v; \lambda_0 + n^{-1/2} w) - \tilde{\mathscr{L}}_n(\theta_0; \lambda_0)$ *and* $\tilde{\mathscr{L}}_n(\theta_0; \lambda_0) = \sum_{t=1}^n \ln f(\tilde{u}_t(\theta_0)\ ;\ \lambda_0)$.

Supplementary appendix for the Maximum likelihood estimation for noninvertible ARMA model with α-stable errors

Additional notation

Let $f_x(x;\lambda) \overset{def}{=} \frac{\partial}{\partial x} f(x;\lambda)$, $f_\lambda(x;\lambda) \overset{def}{=} \frac{\partial}{\partial \lambda} f(x;\lambda)$, $f_{xx}(x;\lambda) \overset{def}{=} \frac{\partial^2}{\partial x^2} f(x;\lambda)$, $f_{\lambda\lambda}(x;\lambda) \overset{def}{=} \frac{\partial^2}{\partial\lambda\partial\lambda^T} f(x;\lambda)$, and $f_{x\lambda}(x;\lambda) \overset{def}{=} \frac{\partial^2}{\partial x \partial \lambda} f(x;\lambda)$. The following notation for the derivatives of the log-densities will be used extensively,

$$
\begin{aligned}
e_x(x_t;\lambda) &\overset{def}{=} \frac{f_x(x_t;\lambda)}{f(x_t;\lambda)}, \quad e_\lambda(x_t;\lambda) \overset{def}{=} \frac{f_\lambda(x_t;\lambda)}{f(x_t;\lambda)}, \\
e_{xx}(x_t;\lambda) &\overset{def}{=} \frac{f_{xx}(x_t;\lambda)}{f(x_t;\lambda)} - \frac{f_x(x_t;\lambda)^2}{f(x_t;\lambda)^2}, \\
e_{x\lambda}(x_t;\lambda) &\overset{def}{=} \frac{f_{\lambda x}(x_t;\lambda)}{f(x_t;\lambda)} - \frac{f_\lambda(x_t;\lambda) d_x(x_t;\lambda)}{f(x_t;\lambda)^2}, \\
e_{xx}(x_t;\lambda) &\overset{def}{=} \frac{f_{xx}(x_t;\lambda)}{f(x_t;\lambda)} - \frac{f_x(x_t;\lambda)^2}{f(x_t;\lambda)^2}, \text{ and} \\
e_{\lambda\lambda}(x_t;\lambda) &\overset{def}{=} \frac{f_{\lambda\lambda}(x_t;\lambda)}{f(x_t;\lambda)} - \frac{f_\lambda(x_t;\lambda) f_\lambda(x_t;\lambda)^T}{f(x_t;\lambda)^2}.
\end{aligned}
$$

Proof of Lemma A2

Proof. $(i) - (vi)$ follows by Lemmas A.1 and A.2. in Meitz and Saikkonen (2013), because they have series presentation in terms of quantities with finite η-moments.

(vii) follows using (3.12), and the fact that $\psi_j^{(b)} \leq \delta^{|j|}$ for some $|\delta| < 1$,

$$
\begin{aligned}
&\mathrm{E}\left[\left(\sum_{t=m_n}^{n-m_n} |\tilde{u}_t(\theta_0) - u_t(\theta_0)|\right)^\eta\right] \leq C \sum_{t=m_n}^{n-m_n} \sum_{j=n-t+1}^{\infty} \delta^{\eta j} \\
=&C \sum_{t=m_n}^{n-m_n} \left(\sum_{j=0}^{\infty} \delta^{\eta j} - \sum_{j=0}^{n-t} \delta^{\eta j}\right) = C \sum_{t=m_n}^{n-m_n} \left(\frac{1}{1-\delta^\eta} - \frac{1-\delta^{\eta(n-t+1)}}{1-\delta^\eta}\right)
\end{aligned}
$$

$$
\begin{aligned}
=&\frac{C}{1-\delta^\eta}\left(\sum_{t=0}^{n_n+1}\delta^{\eta t}-\sum_{t=0}^{n-m_n}\delta^{\eta t}\right)\\
=&\frac{C}{(1-\delta^\eta)^2}\left(\delta^{\eta(n-m_n+1)}-\delta^{\eta(m_n+2)}\right)\to 0,
\end{aligned}
$$

as $n\to\infty$.

In order to show $(viii)$, we show that

$$
\begin{aligned}
&\mathrm{E}\left[\left(\sum_{t=m_n}^{n-m_n}|\tilde{u}_{a_p,t}(\theta_0)-u_{a_p,t}(\theta_0)|\right)^\eta\right]\to 0 \quad \text{and}\\
&\mathrm{E}\left[\left(\sum_{t=m_n}^{n-m_n}|\tilde{u}_{b_q,t}(\theta_0)-u_{b_q,t}(\theta_0)|\right)^\eta\right]\to 0
\end{aligned}
$$

for all $p=1,\ldots,P$ and $q=1,\ldots,Q$, as $n\to\infty$. $(viii)$ follows then, because $v\in\mathbb{R}^{P+Q}$.

Using (3.9) and (3.13),

$$
\begin{aligned}
&\mathrm{E}\left[\left(\sum_{t=m_n}^{n-m_n}|\tilde{u}_{a_p,t}(\theta_0)-u_{a_p,t}(\theta_0)|\right)^\eta\right]\le C\sum_{t=m_n}^{n-m_n}\sum_{j=n-t+1}^{\infty}\delta^{\eta j}\\
=&\frac{C}{1-\delta^\eta}\sum_{t=m_n}^{n-m_n}\delta^{\eta(n-t+1)}=\frac{C}{(1-\delta^\eta)^2}\left(\delta^{\eta(m_n+1)}-\delta^{\eta(n-m_n+2)}\right)\\
&\to 0
\end{aligned}
$$

as $n\to\infty$.

Using (3.9) and (3.13),

$$
\begin{aligned}
&|\tilde{u}_{b_q,t}(\theta_0)-u_{b_q,t}(\theta_0)|\\
\le&\left|\sum_{j=0}^{n-t}\sum_{i=n-t-q-j}^{\infty}\psi_j^{(b)}\psi_i^{(b)}a(B)y_{t+q+i+j}\right|+\left|\sum_{j=n-t+1}^{\infty}\psi_j^{(b)}u_{t+q+j}(\theta_0)\right|.
\end{aligned}
$$

Thus,

$$
\mathrm{E}\left[\left(\sum_{t=m_n}^{n-m_n}|\tilde{u}_{b_q,t}(\theta_0)-u_{b_q,t}(\theta_0)|\right)^\eta\right]
$$

$$
\begin{aligned}
&\leq C\sum_{t=m_n}^{n-m_n}\left(\sum_{j=0}^{n-t}\delta^{\eta j}\sum_{i=n-t-q-j}\delta^{\eta i}+\sum_{j=n-t+1}^{\infty}\delta^{\eta j}\right)\\
&=\frac{C}{1-\delta^{\eta}}\left(\delta^{-\eta(q+1)}\sum_{t=m_n}^{n-m_n}\delta^{\eta(n-t+1)}(n-t+1)+\delta^{\eta(n+m_n+1)}-\delta^{\eta(2n-m_n+2)}\right)\\
&=\frac{C}{1-\delta^{\eta}}\left(\delta^{\eta}\frac{1-(n-m_n+1)\delta^{\eta(n-m_n)}+(n-m_n)\delta^{\eta(n-m_n+1)}}{(1-\delta^{\eta})^2}\right.\\
&\quad\left.-\delta^{\eta}\frac{1-m_n\delta^{\eta m_n}+(m_n-1)\delta^{\eta m_n}}{(1-\delta^{\eta})^2}+\delta^{\eta(n+m_n+1)}-\delta^{\eta(2n-m_n+2)}\right)\\
&\quad\to 0
\end{aligned}
$$

as $n\to\infty$.

□

Proofs for the Asymptotic results

Proof of Lemma B3

Step I For $v\in\mathbb{R}^{P+Q}$ and $w\in\mathbb{R}^4$,

$$
\begin{aligned}
&\sum_{t=1}^{n}\ln f(u_t(\theta_0+n^{-1/\alpha_0}v)\ ;\ \lambda_0+n^{-1/2}w)\\
&-\sum_{t=1}^{n}\ln f(\varepsilon_t+n^{-1/\alpha_0}v^T u_{\theta,t}(\theta_0)\ ;\ \lambda_0+n^{-1/2}w)=o_p(1)
\end{aligned}
$$

as $n\to\infty$.

Proof. Using mean vale theorem, the previous equals to

$$
\begin{aligned}
&\sum_{t=1}^{n}e_x(x_{t,n}^*(v)\ ;\ \lambda_0+n^{-1/2}w)\times[u_t(\theta_0+n^{-1/\alpha_0}v)-\varepsilon_t-n^{-1/\alpha_0}v^T u_{\theta,t}(\theta_0)]\\
=&\sum_{t=1}^{n}e_x(x_{t,n}^*(v)\ ;\ \lambda_0+n^{-1/2}w)\times n^{-2/\alpha_0}v^T u_{\theta\theta,t}(\theta_n^*(v))v
\end{aligned}
$$

with $x_{t,n}^*(v)$ between $u_t(\theta_0+n^{-1/\alpha_0}v)$ and $\varepsilon_t+n^{-1/\alpha_0}v^T u_{\theta,t}(\theta_0)$, and $\theta_n^*(v)$ between θ_0 and $\theta_0+n^{-1/\alpha_0}v$. Let us consider values of v and w inside a δ-

ball d_δ, such that for $v \in d_{\delta,v}$ and $v \in d_{\delta,w}$, we have $\|v\| < (P+Q)^{1/2}\delta$ and $\|w\| < 2\delta$. For $v \in d_{\delta,v}$ and $w \in d_{\delta,w}$, the previous can be bounded above by

$$\sup_{v\in d_{\delta,v}, w\in d_{\delta,w}} \left| e(x^*_{t,n}(v); \lambda_0 + n^{-1/2}w) \right| \times n^{-2/\alpha_0}(P+Q)^2\delta^2 \sum_{t=1}^{n} \sup_{v\in d_{\delta,v}} |u_{\theta\theta,t}(\theta^*_n(v))| .$$

The first part is finite since $f(\cdot;\cdot)$ is bounded and continuously differentiable. The second part can be evaluated with Markov's inequality as

$$\begin{aligned} & P\left(\left| n^{-2/\alpha_0}(P+Q)^2\delta^2 \sum_{t=1}^{n} \sup_{v\in d_{\delta,v}} |u_{\theta\theta,t}(\theta^*_n(v))| \right| > \epsilon \right) \\ \overset{0\le\kappa_1\le 1}{\le} & \left(\frac{n^{-2/\alpha_0}(P+Q)^2\delta^2}{\epsilon} \right)^{\kappa_1} \mathrm{E}\left[\left| \sum_{t=1}^{n} \sup_{v\in d_{\delta,v}} |u_{\theta\theta,t}(\theta^*_n(v))| \right|^{\kappa_1} \right] \\ \le & \left(\frac{n^{-2/\alpha_0}(P+Q)^2\delta^2}{\epsilon} \right)^{\kappa_1} n\mathrm{E}\left[\left| \sum_{t=1}^{n} \sup_{v\in d_{\delta,v}} |u_{\theta\theta,t}(\theta^*_n(v))| \right|^{\kappa_1} \right] \\ = & n^{1-2\kappa_1/\alpha_0} \left(\frac{(P+Q)^2\delta^2}{\epsilon} \right)^{\kappa_1} \mathrm{E}\left[\left| \sum_{t=1}^{n} \sup_{v\in d_{\delta,v}} |u_{\theta\theta,t}(\theta^*_n(v))| \right|^{\kappa_1} \right]. \end{aligned}$$

Provided by Lemma A1, there is $\alpha_0/2 < \kappa_1 < 1$ such that the the latter part is finite. For any small ϵ, let $n \to \infty$, and the result follows. □

Step II For $v \in \mathbb{R}^{P+Q}$ and $w \in \mathbb{R}^4$,

$$\begin{aligned} & \sum_{t=1}^{n} \ln f(\varepsilon_t + n^{-1/\alpha_0} v^T u_{\theta,t}(\theta_0) \; ; \; \lambda_0 + n^{-1/2}w) \\ & - \sum_{t=1}^{n} \ln f(\varepsilon_t + n^{-1/\alpha_0} v^T u_{\theta,t}(\theta_0) \; ; \; \lambda_0) \\ & - n^{-1/2} w^T \sum_{t=1}^{n} e_\lambda(\varepsilon_t; \lambda_0) + \frac{1}{2} w^T \mathbf{I}(\lambda_0) w = o_p(1). \end{aligned}$$

Proof. Approximating the expression with 2^{nd} order Taylor approximation gives

$$n^{-1/2}w^T\left[\sum_{t=1}^{n} e_{\lambda}(\varepsilon_t + n^{-1/\alpha_0}v^T u_{\theta,t}(\theta_0)\ ;\ \lambda_0) - \sum_{t=1}^{n} e_{\lambda,t}(\varepsilon_t\ ;\ \lambda_0)\right]$$
$$+\frac{1}{2n}w^T\sum_{t=1}^{n} e_{\lambda\lambda}(\varepsilon_t + n^{-1/\alpha_0}v^T u_{\theta,t}(\theta_0)\ ;\ \lambda_n^*(w))w + \frac{1}{2}w^T\mathbf{I}(\lambda_0)w.$$

Second row is clearly $o_p(1)$ by the ergodic theorem and by the fact that $f(\cdot;\cdot)$ is twice continuously differentiable. To see that the first row is also $o_p(1)$, use the mean value expansion to get

$$n^{-1/2-1/\alpha_0}w^T\sum_{t=1}^{n} e_{x\lambda}(x_{n,t}^{**}(v)\ ;\ \lambda_0)v^T u_{\theta,t}(\theta_0)$$

with $x_{n,t}^{**}(v)$ between ε_t and $\varepsilon_t + n^{-1/\alpha_0}v^T u_{\theta,t}(\theta_0)$. This can be bounded from above by

$$n^{-1/2-1/\alpha_0}\sup_{w\in d_\delta, x\in\mathbf{R}}\left|w^T e_{x\lambda}(x;\lambda_0)\right|\sum_{t=1}^{n}\sup_{v\in d_\delta}\left|v^T u_{\theta,t}(\theta_0)\right|$$

The term in the middle is bounded because because $f(\cdot\ ;\cdot)$ is twice continuously differentiable. Using (3.11), the last part is

$$\sum_{t=1}^{n}\sup_{v\in d_\delta}\left|\sum_{j=-\infty}^{\infty} c_j(v)\varepsilon_{t-j}\right| \le \sum_{t=1}^{n}\sum_{j=-\infty}^{\infty} K_1 c^{|j|}|\varepsilon_{t-j}|.$$

Evaluate the probabilities by Markov's inequality,

$$P\left(\left|n^{-1/2-1/\alpha_0}\sup_{w\in d_\delta, x\in\mathbb{R}}\left|w^T e_{\lambda x}\left(x;\lambda_0\right)\right|\sum_{t=1}^{n}\sum_{j=-\infty}^{\infty} K_1 c^{|j|}|\varepsilon_{t-j}||\right| > \epsilon\right)$$
$$\overset{|\kappa_2|\le 1}{\le}\left(\frac{n^{-1/2-1/\alpha_0}\sup_{w\in d_\delta, x\in\mathbb{R}}\left|w^T e_{\lambda x}(x;\lambda_0)\right|}{\epsilon}\right)^{\kappa_2}\mathrm{E}\left[\left|\sum_{t=1}^{n}\sum_{j=-\infty}^{\infty} K_1 c^{|j|}|\varepsilon_{t-j}|\right|^{\kappa_2}\right]$$
$$\le\left(\frac{n^{-1/2-1/\alpha_0}\sup_{w\in d_\delta, x\in\mathbb{R}}\left|w^T e_{\lambda x}(x;\lambda_0)\right|}{\epsilon}\right)^{\kappa_2}|K_1|^{\kappa_2}\sum_{t=1}^{n}\sum_{j=-\infty}^{\infty} c^{|j|}\mathrm{E}\left[|\varepsilon_t|^{\kappa_2}\right]$$

$$\leq \left(\frac{n^{-1/2-1/\alpha_0} \sup_{w \in d_\delta, x \in \mathbb{R}} \left| w^T e_{\lambda x}(x;\lambda_0) \right|}{\epsilon} \right)^{\kappa_2} |K_1|^{\kappa_2} n \mathrm{E}\left[|\varepsilon_t|^{\kappa_2} \right] \sum_{j=-\infty}^{\infty} \left| c^{|j|} \right|^{\kappa_2}.$$

The result follows by choosing κ_2 such that

$$\kappa_2 = \begin{cases} 1 & \text{if } \alpha_0 \geq 1 \\ \frac{\alpha_0}{2} + \frac{\alpha_0}{2+\alpha_0} & \text{if } \alpha_0 < 1 \end{cases}.$$

This way $|\kappa_2| < 1$, $\mathrm{E}[|\varepsilon|^{\kappa_2}] < \infty$, and $n^{1-(1/2-1/\alpha_0)\kappa_2} \to 0$ as $n \to \infty$. $\square$

Step III For $v \in \mathbb{R}^{P+Q}$ and $w \in \mathbb{R}^4$, and

$$\begin{aligned} Q_n^*(v) = &\sum_{t=1}^{n} [\ln f(u_t(\theta_0 + n^{-1/\alpha_0} v^T u_{\theta,t}(\theta_0)) \ ; \ \lambda_0) - \ln f(\varepsilon_t \ ; \ \lambda_0)] \\ &+ n^{-1/2} w^T \sum_{t=1}^{n} e_\lambda(\varepsilon_t;\lambda_0) \end{aligned}$$

we have

$$\mathcal{Q}_n(v,w) - \mathcal{Q}_n^*(v) = w^T \mathbf{I}(\lambda_0) w + o_p(1).$$

We obtain the result by re-writing the left hand side by adding and subtracting a term, and then using Step I and Step II,

$$\begin{aligned} \mathcal{Q}_n(v,w) - \mathcal{Q}_n^*(v) = &\sum_{t=1}^{n} \ln f(u_t(\theta_0 + n^{-1/\alpha_0} v^T u_{\theta,t}(\theta_0)) \ ; \ \lambda_0 + n^{-1/2} w) \\ &+ \sum_{t=1}^{n} \ln f(\varepsilon_t + n^{-1/\alpha_0} v^T u_{\theta,t}(\theta_0) \ ; \ \lambda_0 + n^{-1/2} w) \\ &- \sum_{t=1}^{n} \ln f(\varepsilon_t + n^{-1/\alpha_0} v^T u_{\theta,t}(\theta_0) \ ; \ \lambda_0 + n^{-1/2} w) \\ &- \sum_{t=1}^{n} \ln f(\varepsilon_t + n^{-1/\alpha_0} v^T u_{\theta,t}(\theta_0) \ ; \ \lambda_0) \\ &- n^{-1/2} w^T \sum_{t=1}^{n} e_\lambda(\varepsilon_t;\lambda_0) \end{aligned}$$

$$
\begin{aligned}
&= \sum_{t=1}^{n} \ln f(\varepsilon_t + n^{-1/\alpha_0} v^T u_{\theta,t}(\theta_0)\ ;\ \lambda_0 + n^{-1/2} w) \\
&\quad - \sum_{t=1}^{n} \ln f(\varepsilon_t + n^{-1/\alpha_0} v^T u_{\theta,t}(\theta_0)\ ;\ \lambda_0) \\
&\quad - n^{-1/2} w^T \sum_{t=1}^{n} e_\lambda(\varepsilon_t; \lambda_0) + o_p(1) \\
&= \frac{1}{2} w^T \mathbf{I}(\lambda_0) w + o_p(1).
\end{aligned}
$$

Proof of Lemma B4

The result follows by Theorem 3 on p. 37 of Rosenblatt (1985), by showing that for all the cumulants $\mathrm{cum}_k(\gamma_1 \mathcal{W}_{m,n}(A_s) + \gamma_2 N_n(w)) \to \mathrm{cum}_k(\gamma_1 \mathcal{W}_m(A_s) + \gamma_2 w^T \mathbf{N})$, where $\mathrm{cum}_k(X)$ is the k^{th} order cumulant $\mathrm{cum}(X, \ldots, X)$ (k times X's). If this holds for all $\gamma_1, \gamma_2 \in \mathbb{R}$, then by Cramér-Wold theorem $(\mathcal{W}_{m,n}(A_s), N_n(w)) \xrightarrow{d} (\mathcal{W}_m(A_s), w^T \mathbf{N})$.

Begin by noticing that

$$
\begin{aligned}
\mathrm{cum}_k(\gamma_1 \mathcal{W}_{m,n}(A) + \gamma_2 N_n(w)) =& \gamma_1^k \mathrm{cum}_k(\mathcal{W}_{m,n}(A)) + \gamma_2^k \mathrm{cum}_k(N_n(w)) \\
&+ \sum_{j=1}^{k-1} \binom{k}{j} \gamma_1^j \gamma_2^k \mathrm{cum}_{j,k-j}\left(\mathcal{W}_{m,n}(A), N_n(w)\right),
\end{aligned}
$$

where

$$
\begin{aligned}
&\mathrm{cum}_{j,k-j}\left(\sum_{t=1}^{n} I_{\left(\varepsilon_t, c(\alpha_0)^{-1}\sigma_0^{-1} n^{-1/\alpha_0} \mathbf{e}_t\right)}(A), n^{-1/2} \sum_{t=1}^{n} w^T e_\lambda\left(\varepsilon_t\ ;\ \lambda_0\right)\right) \\
= &\sum_{t_1=1}^{n} \cdots \sum_{t_k=1}^{n} n^{-\frac{k-j}{2}} \mathrm{cum}(\phi_{t_1,n}(A), \ldots, \phi_{t_j,n}(A), \varphi_{t_{j+1}}(w), \ldots, \varphi_{t_k}(w)) \quad (3.18)
\end{aligned}
$$

with $\phi_{t,n}(A) = I_{\left(\varepsilon_t, c(\alpha_0)^{-1}\sigma_0^{-1} n^{-1/\alpha_0} \mathbf{e}_t\right)}(A)$ and $\varphi_t(w) = w^T e_\lambda(\varepsilon_t; \lambda_0)$.

Step I: (3.18) **converges to zero as** $n \to \infty$

The next step makes use of the dependence structure of these random variables. Independence of $\phi_{t,n}(A)$ and $\varphi_{t+j}(w)$ is clear for all $|j| > m$. Also, $\phi_{t,n}(A)$ and $\phi_{t+s,n}(A)$ are independent for all $|s| > 2m$.

Fix t_1 and consider $\phi_{t_s,n}(A)$ for any $s = 2, \ldots, j$, such that $|t_1 - t_s| > 2m(j-1)$. For these indices, $\phi_{t_1,n}(A)$ and $\phi_{t_s,n}(A)$ are always independent and we can find a subset of random variables from $(\phi_{t_1,n}(A), \ldots, \phi_{t_j,n}(A))$ that is independent of the rest of that set. In expression (3.18) the summands where one of the indices satisfy $|t_1 - t_s| > 2m(j-1)$ can be pruned since they do not contribute to the sum. Using the same logic, we can extend this idea also to the indices t_s with $s = (j+1), \ldots, k$. If there is a subsets in the set $(\phi_{t_1,n}(A), \ldots, \phi_{t_j,n}(A))$ that is independent of every other subsets, $\varphi_{t_l}(w)$, $l = j+1, \ldots, k$, can be dependent only with that particular subset or some other subsets, never both. Then again, if $|t_1 - t_l| > m(2j-1)$ for some $l = j+1, \ldots, k$, we can apply the previous reasoning again.

We have an equivalent expression for the sum in (3.18) as

$$\sum_{t_1=1}^{n} \sum_{|t_1-t_2|\le 2m(j-1)} \cdots \sum_{|t_1-t_j|\le 2m(j-1)} \sum_{|t_1-t_{j+1}|\le m(2j-1)} \cdots \sum_{|t_1-t_k|\le m(2j-1)}$$
$$n^{-\frac{k-j}{2}} \operatorname{cum}(\phi_{t_1,n}(A), \ldots, \phi_{t_j,n}(A), \varphi_{t_{j+1}}(w), \ldots, \varphi_{t_k}(w)). \tag{3.19}$$

Next we show that the expression (3.19) is $o(1)$. To this end, notice that the expression (3.19) is made up of $n(4m(j-1)+1)^{j-1}(2m(2j-1)+1)^{k-j}$ terms. Note, that the cumulant in (3.19) is always finite since $|\phi_{t_s,n}(A)| \le 1$ and $\mathrm{E}[|\varphi_{t_l}(w)|^r] < \infty$ for all $s = 1, \ldots, j$, $l = j+1, \ldots, k$ and $r < \infty$. As the number of terms in the sum increases linearly with n, scaling by $n^{-\frac{k-j}{2}}$ is enough to ensure the convergence as long as $k - j \ge 3$.

We show the same result for $k - j < 3$ in parts. This means, that in the cumulant above there are up to two $\varphi_{t_l}(w)$ terms. The cumulant in expression (3.19) consists of terms like

(i) $\mathrm{E}\left[\phi_{t_1,n}(A)\varphi_{t_2}(w)\right]$ and

(ii) $\mathrm{E}\left[\phi_{t_{1,n}}(A)\varphi_{t_2}(w)\varphi_{t_3}(w)\right]$,

terms with more coefficients $\phi_{t_l,n}(A)$, and their products. Terms with more than one $\phi_{t_l,n}(A)$ coefficients have expectation less or equal than that in (ii), so we do not have to handle those explicitly. We show that $n^{1/2}$ times the terms (i) and (ii) are $o(1)$ for all indices, which ensures that expression (3.19) is $o(1)$.

Let us begin by showing this for (i). Using the continuous differentiability of the density function and Lemma 1 (*ii*), (*iv*) and (*v*), we can find constants C_e and D_e such that

$$\left|w^T e_\lambda(\varepsilon_t;\lambda_0)\right| \leq C_e + D_e|x|^{\alpha_0/4}, \quad \text{for all } x \in \mathbb{R}.$$

Using this finding we can bound (i) by

$$\begin{aligned} &C_e\mathrm{E}\left[I_{\left(\varepsilon_{t_1},c(\alpha_0)^{-1}\sigma_0^{-1}n^{-1/\alpha_0}\mathbf{e}_{t_1}\right)}(A)\right] + \\ &D_e\mathrm{E}\left[|\varepsilon_{t_2}|^{\alpha_0/4} I_{\left(\varepsilon_{t_1},c(\alpha_0)^{-1}\sigma_0^{-1}n^{-1/\alpha_0}\mathbf{e}_{t_1}\right)}(A)\right]. \end{aligned} \tag{3.20}$$

Recall that the origin is not included in the set $A = (\underline{a}_{-m}, \overline{a}_{-m}] \times \cdots \times (\underline{a}_m, \overline{a}_m]$. Assume that $(\underline{a}_1, \overline{a}_1]$ does not contain zero, the index does not matter here. Expression (3.20) can be further bounded above by

$$C_e P\left(\varepsilon_{t_1+1} \geq \underline{a}_1 c(\alpha_0)\sigma_0 n^{1/\alpha_0}\right) + D_e\mathrm{E}\left[|\varepsilon_{t_2}|^{\alpha_0/4} I\left\{\varepsilon_{t_1+1} \geq \underline{a}_1 c(\alpha_0)\sigma_0 n^{1/\alpha_0}\right\}\right].$$

The first term is $o(n)$ and Karamata's theorem[6] (Feller, 1971, p. 283) implies that the second part has the same limit as

$$(\text{const}) \times D_e(\underline{a}_1 c(\alpha_0)\sigma_0 n^{1/\alpha_0})^{\alpha_0/4} P\left(|\varepsilon_{t'}| \geq \underline{a}_1 c(\alpha_0)\sigma_0 n^{1/\alpha_0}\right) = o(n^{3/4}).$$

Multiplied by $n^{1/2}$, these terms still converges to zero.

Very similar steps can be used to establish (ii) as well.

It has been shown that

$$\mathrm{cum}_k(\gamma_1 \mathcal{W}_{m,n}(A) + \gamma_2 N_n(w)) \to \gamma_1^k \mathrm{cum}_k(\mathcal{W}_{m,n}(A)) + \gamma_2^k \mathrm{cum}_k(N_n(w)).$$

[6]See Stochastic Models with Power-Law Tails: The Equation X = AX + B by Mikosch et al., Appendix B3.

Step II: $\mathbf{cum}_k\left(\mathcal{W}_{m,n}(A)\right) \to \mathbf{cum}_k\left(\mathcal{W}_m(A)\right)$ **as** $n \to \infty$

Using previous arguments, we have seen that

$$\begin{aligned}\operatorname{cum}_k(\mathcal{W}_{m,n}(A)) &= \sum_{t_1=1}^{n} \cdots \sum_{t_k=1}^{n} \operatorname{cum}(\phi_{t_1,n}(A), \ldots, \phi_{t_k,n}(A)) \\ &= \sum_{t_1=1}^{n} \sum_{|t_1-t_2|\le 2m(k-1)} \cdots \sum_{|t_1-t_k|\le 2m(k-1)} \\ &\quad \operatorname{cum}(\phi_{t_1,n}(A), \ldots, \phi_{t_k,n}(A)). \end{aligned} \tag{3.21}$$

From the page 42 in Rosenblatt (2012) we have

$$\begin{aligned}&\operatorname{cum}(\phi_{t_1,n}(A), \ldots, \phi_{t_k,n}(A)) \\ &= \sum (-1)^{p-1}(p-1)!\, \mathrm{E}\left[\prod_{j\in\nu_1} \phi_{j,n}(A)\right] \cdots \mathrm{E}\left[\prod_{j\in\nu_p} \phi_{j,n}(A)\right],\end{aligned}$$

with $\nu_1, \ldots, \nu_p$ one particular partition of $(t_1, \ldots, t_k)$ and the summation is over all possible partitions (number of partitions depends only on k, not n). For our purpose, in each partition, there are two different classes of subsets ν_j: one where all the indices in ν_j are the same, and one with at least one index different from the rest. In the former case, say $\nu_j = (1, \ldots, 1)$,

$$\mathrm{E}\left[\phi_{1,n}(A)^s\right] = \mathrm{E}\left[\phi_{1,n}(A)\right] < \infty.$$

For the latter case, consider, say $\nu_j = (1,1,2,2,2)$. From the definition of the set A in Lemma B4, fix the interval that does not contain zero, say $(\underline{a}_1, \bar{a}_1]$. We get

$$\begin{aligned}&n\mathrm{E}\left[\phi_{1,n}(A)^2\phi_{2,n}(A)^3\right] = n\mathrm{E}\left[\phi_{1,n}(A)\phi_{2,n}(A)\right] \\ &\le nP\left((|c(\alpha_0)^{-1}\sigma_0^{-1}n^{-1/\alpha_0}\varepsilon_2| > \underline{a}_1) \bigcap (|c(\alpha_0)^{-1}\sigma_0^{-1}n^{-1/\alpha_0}\varepsilon_3| > \underline{a}_1)\right) \to 0,\end{aligned}$$

as $n \to \infty$. Same logic can be used to obtain this result for all the terms $\mathrm{E}\left[\prod_{j\in\nu_j} \phi_{j,n}(A)\right]$ where the subset ν_j consists of at least two different indices.

There are exactly n summands in (3.21) where all the indices $(t_1, \ldots, t_k)$ coincide. Also the number of the rest of the summands is growing linearly in

n, so we can write (3.21) as

$$\mathrm{cum}_k(W_{m,n}(A)) = n\ \mathrm{cum}_k(\phi_{1,n}(A)) + \text{resid.}$$

with resid. $\to 0$ as $n \to 0$. Using the fact that

$$n\mathrm{E}[\phi_{1,n}(A)]\mathrm{E}[\phi_{1,n}(A)] \to 0$$

as $n \to \infty$, it boils down to showing that

$$nP\left(\varepsilon_1, c(\alpha_0)^{-1}\sigma_0^{-1}n^{-1/\alpha_0}\mathbf{e}_1 \in A\right) \to \mu(A)$$

where $\mu(A)$ is a mean measure of Poisson random variable $\mathcal{W}_m(A)$ for a fixed A (see section 3.4. in Calder (1998)).

Step III: $\mathbf{cum}_k(N_n(w)) \to \mathbf{cum}_k(w^T\mathbf{N})$ as $n \to \infty$

Notice that

$$\mathrm{cum}_k\left(n^{-1/2}\sum_{t=1}^{n} e_\lambda(\varepsilon_t;\lambda_0)\right) = n^{1-k/2}\mathrm{cum}_k(w^T e_\lambda(\varepsilon_t;\lambda_0)). \tag{3.22}$$

For $k > 2$ this converges to zero. For $k = 2$ (3.22) is

$$\mathrm{E}[w^T e_\lambda(\varepsilon_t;\lambda_0)e_{\lambda^T}(\varepsilon_t;\lambda_0)w] = w^T\mathbf{I}(\lambda_0)w.$$

For $k = 1$, (3.18) is zero. We have established cumulants of a normally distributed random variable $w^T\mathbf{N} \sim N(0, w^T\mathbf{I}(\lambda_0)w)$.

Steps I-III establish that $(\mathcal{W}_{m,n}(A), N_n(w)) \xrightarrow{d} (\mathcal{W}_n(A), w^T\mathbf{N})$ on $\mathbb{R}^2$ for all A defined in Lemma B4. Same arguments can be used to generalize this result to a three dimensional vector $(\mathcal{W}_{m,n}(A_1), \mathcal{W}_{m,n}(A_2), N_n(w))$, and iterativaly to $(l+1)$-dimensional vector in Lemma B4. To this end, notice that

$$\begin{aligned}
&\mathrm{cum}_k(\gamma_1\mathcal{W}_{m,n}(A_1) + \gamma_2\mathcal{W}_{m,n}(A_2) + \gamma_3 N_n(w)) \\
=&\gamma_1^k\mathrm{cum}_k(\mathcal{W}_{n,n}(A_1)) + \mathrm{cum}_k(\gamma_2 W_{m,n}(A_2) + \gamma_3 N_n(w)) && (3.23)\\
&+\sum_{j=1}^{k-1}\binom{k}{j}\mathrm{cum}_{j,k-j}(\gamma_1\mathcal{W}m,n(A_1), \gamma_2\mathcal{W}_{m,n}(A_2) + \gamma_3 N_n(w)). && (3.24)
\end{aligned}$$

We have already established the limit of the first row (3.23), which is

$$\mathrm{cum}_k(\mathcal{W}_m(A_1), \mathcal{W}_m(A_2), w^T\mathbf{N}).$$

For (3.24), we write

$$\begin{aligned}
&\mathrm{cum}_{j,k-j}(\gamma_1\mathcal{W}_{m,n}(A_1), \gamma_2\mathcal{W}_{m,n}(A_2) + \gamma_3 N_n(w)) \\
=&\sum_{l=0}^{k-j}\binom{k-j}{l}\mathrm{cum}_{j,l,k-j-l}(\gamma_1\mathcal{W}_{m,n}(A_1), \gamma_2\mathcal{W}_{m,n}(A_2), \gamma_3 N_n(w)) \\
=&\sum_{l=0}^{k-j}\binom{k-j}{l} n^{-\frac{k-j-l}{2}}\gamma_1^j\gamma_2^l\gamma_3^{k-j-l}\sum_{t_1=1}^{n}\cdots\sum_{t_k=1}^{n} \\
&\mathrm{cum}(\phi_{t_1,n}(A_1), \ldots, \phi_{t_j,n}(A_1), \phi_{t_{j+1},n}(A_2), \ldots \\
&\ldots, \phi_{t_{j+l},n}(A_2), \varphi_{t_{j+l+1},n}(w), \ldots, \varphi_{t_k,n}(w)).
\end{aligned}$$

Previous arguments can be applied to show that this converges to zero as $n \to \infty$, and furthermore that (3.24) converges to zero, and thus, Lemma B4 holds for any dimension l.

□

Proof of Lemma B5

We begin by repeating the arguments in Calder (1998), Lemmas 5-8, which provides the proof relying on the symmetry arguments.

We begin by repeating the arguments in Calder (1998), Lemmas 5-8 to show the result for $0 < \alpha_0 < 2$, relying on the symmetry arguments of the density function. After that we show that for $0 < \alpha_0 < 1$ the result follows without any assumption on the symmetry.

Let us denote $K_{k,j} = c(\alpha_0)^{1/\alpha_0}\sigma_0 c_j(v)\delta_k\Gamma_k^{-1/\alpha_0}$, and $I_{t,n}^{M,\delta} = I\{|\varepsilon_{k,j}| \leq M\}I\{|K_{k,j}| \geq \delta\}$. Note that $1 - I_{t,n}^{M,delta} = I\{|K_{k,j}| \leq \delta\} + I\{|\varepsilon_{k,j}| > M\}I\{|K_{k,j}| > \delta\} \overset{def.}{=} I_{k,j}^{A,\delta} + I_{k,j}^{B,M}$, with M and δ some possibly large and small positive real numbers, respectively.

We show that

$$P\left(\sum_{k=1}^{\infty}\sum_{j=1}^{\infty}\left|\ln f(\varepsilon_{k,j} + K_{k,j}\ ;\ \lambda_0) - \ln f(\varepsilon_{k,j};\lambda_0)\right| \times (1 - I_{k,j}^{M,\delta}) > \epsilon\right) \quad (3.25)$$

$$\to 0$$

as $M \to \infty$ and $\delta \to 0$. Using Taylor's approximation we get

$$P\left(\sum_{k=1}^{\infty}\sum_{j=1}^{\infty}\left|e_x(\varepsilon_{k,j};\lambda_0)K_{k,j} + (e_x(\xi_{k,j};\lambda_0) - e_x(\varepsilon_{k,j};\lambda_0))K_{k,j}\right| \times (1 - I_{k,j}^{M,\delta}) > \epsilon\right)$$

$$\begin{aligned}
\leq &P\left(\sum_{k=1}^{\infty}\sum_{j=1}^{\infty}\left|e_x(\varepsilon_{k,j};\lambda_0)K_{k,j}\right| \times (1 - k_{k,j}^{M,\delta}) > \frac{\epsilon}{2}\right)\\
&+ P\left(\sum_{k=1}^{\infty}\sum_{j=1}^{\infty}\left|(e_x(\xi_{k,j};\lambda_0) - e_x(\varepsilon_{k,j};\lambda_0))K_{k,j}\right| \times (1 - k_{k,j}^{M,\delta}) > \frac{\epsilon}{2}\right)
\end{aligned}$$

with $|\varepsilon_{k,j} - \xi_{k,j}| \leq K_{k,j}$. Using this, together with continuity of $e_x(\cdot;\lambda_0)$, and the partition of $1 - I_{k,j}^{M,\delta}$, we get

$$\begin{aligned}
\leq &P\left(\sum_{k=1}^{\infty}\sum_{j=1}^{\infty}\left|e_x(\varepsilon_{k,j};\lambda_0)K_{k,j}\right| \times I_{k,j}^{A,\delta} > \frac{\epsilon}{4}\right)\\
&+ P\left(\sum_{k=1}^{\infty}\sum_{j=1}^{\infty}\left|e_x(\varepsilon_{k,j};\lambda_0)K_{k,j}\right| \times I_{k,j}^{B,M,\delta} > \frac{\epsilon}{4}\right)\\
&+ P\left(\sum_{k=1}^{\infty}\sum_{j=1}^{\infty}\left|K_{k,j}\right|^2 \times I_{k,j}^{A,\delta} > \frac{\epsilon}{4}\right) + P\left(\sum_{k=1}^{\infty}\sum_{j=1}^{\infty}\left|e_x(\varepsilon_{k,j};\lambda_0)K_{k,j}\right| \times I_{k,j}^{B,M,\delta} > \frac{\epsilon}{4}\right).
\end{aligned}$$

Each of these terms go to zero as $M \to \infty$ and $\delta \to 0$, by Lemmas 5-8 in Calder (1998), provided that $\mathrm{E}[e_x(\varepsilon_t;\lambda_0)] = 0$.

Next we look at the situation $0 < \alpha_0 < 1$. The absolute value in (3.25) can be approximated by

$$K_{k,j}e_x(\xi_{k,j};\lambda_0) \leq K_{k,j}\sup_{x\in\mathbb{R}}|e_x(x;\lambda_0)|,$$

where $\xi_{k,j}$ is between $\varepsilon_{k,j}$ and $\varepsilon_{k,j} + K_{k,j}$. For $0 < \alpha < 1$, $\sum_{n=1}^{\infty}\Gamma_k^{-1/\alpha_0} < \infty$ a.s. (Samoradnitsky and Taqqu, 1994, p. 30). Using this, and the geometric convergence of the sequence $c_j(v)$,

$$c_0(\alpha)^{1/\alpha_0}\sigma_0 \sup_{x\in\mathbb{R}} |e_x(x;\lambda_0)| \sum_{j=1}^{\infty} |c_j(v)| \sum_{k=1}^{\infty} |\Gamma_k^{-1/\alpha_0}| < \infty \quad \text{a.s..}$$

This gives the convergence in (3.25) as $\delta \to 0$.

□

Proof of Lemma B6

This proof is similar to the proof of Lemma A.10. in Andrews et al. (2009). In order to show the result, we need results like those in Calder (1998), Lemmas 1-4 and Proposition A.2. (a)-(c) in Davis et al. (1992).

Let $X_t^+(v) \overset{def}{=} \sum_{j=1}^{\infty} c_{-j}(v)\varepsilon_{t+j}$ and $X_t^-(v) \overset{def}{=} \sum_{j=1}^{\infty} c_j(v)\varepsilon_{t-j}$. Note, that $X_t^+(v)$ and $X_t^-(v)$ are α-stable. Let $U_{t,n}^+(v) \overset{def}{=} n^{-1/\alpha_0} X_t^+(v)$ and $U_{t,n}^-(v) \overset{def}{=} n^{-1/\alpha_0} X_t^-(v)$. Now, $X_t^+(v)$ is independent of $X_t^-(v)$, and thus $U_{t,n}^+(v)$ is independent of $U_{t,n}^-(v)$. For the sake of brevity, for now on, we will not write $U_{t,n}^{\pm}$ as a function of v, explicitly. Let $I_{t,n}^{M,\delta} \overset{def}{=} I\{|\varepsilon_t| < MI\{|U_{t,n}^+| > \delta \cup |U_{t,n}^-| > \delta\}$. Using the complements of these events, we have

$$\begin{aligned}1 - I_{t,n}^{M,\delta} =& I\{|U_{t,n}^+| < \delta\}I\{|U_{t,n}^-| < \delta\} + I\{|\varepsilon_t| > M\}I\{|U_{t,n}^+| > \delta\} \\ &+ I\{|\varepsilon| > M\}I\{|U^-| > \delta\} - I\{|\varepsilon_t| > M\}I\{|U_{t,n}^+| > \delta\}I\{|U^-| > \delta\} \\ \overset{def}{=}& I_{t,n}^{A,\delta} + I_{t,n}^{B,M,\delta} + I_{t,n}^{C,M,\delta} - I_{t,n}^{D,M,\delta}. \qquad (3.26)\end{aligned}$$

Using 2^{nd} order Taylor approximation, we need to show that

$$\begin{aligned}&\lim_{M\to\infty}\lim_{\delta\to 0}\limsup_{n\to\infty} \\ &P\left(\left|\sum_{t=1}^{n}[(U_{t,n}^+ + U_{t,n}^-)e_x(\varepsilon_t;\lambda_0) + \frac{1}{2}(U_{t,n}^+ + U_{t,n}^-)^2 e_{xx}(x_t^*;\lambda_0)]\right| \right. \qquad (3.27) \\ &\left. \times(1 - I_{t,n}^{M,\delta}) > \epsilon\right) = 0\end{aligned}$$

with $x_{t,n}^*$ between ε_t and $\varepsilon_t + U_{t,n}^+ + U_{t,n}^-$. Using additivity of probability and expression (3.42), the probability in (3.43) can be bounded from above by

$$P\left(\left|\sum_{t=1}^{n}(U_{t,n}^{+}+U_{t,n}^{-})e_x(\varepsilon_t;\lambda_0)\right|\times I_{t,n}^{A,\delta}>\frac{1}{8}\epsilon\right) \tag{3.28}$$

$$+P\left(\left|\sum_{t=1}^{n}(U_{t,n}^{+}+U_{t,n}^{-})e_x(\varepsilon_t;\lambda_0)\right|\times I_{t,n}^{B,M,\delta}>\frac{1}{8}\epsilon\right) \tag{3.29}$$

$$+P\left(\left|\sum_{t=1}^{n}(U_{t,n}^{+}+U_{t,n}^{-})e_x(\varepsilon_t;\lambda_0)\right|\times I_{t,n}^{C,M,\delta}>\frac{1}{8}\epsilon\right) \tag{3.30}$$

$$+P\left(\left|\sum_{t=1}^{n}(U_{t,n}^{+}+U_{t,n}^{-})e_x(\varepsilon_t;\lambda_0)\right|\times I_{t,n}^{D,M,\delta}>\frac{1}{8}\epsilon\right) \tag{3.31}$$

$$+P\left(\left|\sum_{t=1}^{n}\frac{1}{2}(U_{t,n}^{+}+U_{t,n}^{-})^2e_{xx}(x_{t,n}^{*};\lambda_0)\right|\times I_{t,n}^{A,\delta}>\frac{1}{8}\epsilon\right) \tag{3.32}$$

$$+P\left(\left|\sum_{t=1}^{n}\frac{1}{2}(U_{t,n}^{+}+U_{t,n}^{-})^2e_{xx}(x_{t,n}^{*};\lambda_0)\right|\times I_{t,n}^{B,M,\delta}>\frac{1}{8}\epsilon\right) \tag{3.33}$$

$$+P\left(\left|\sum_{t=1}^{n}\frac{1}{2}(U_{t,n}^{+}+U_{t,n}^{-})^2e_{xx}(x_{t,n}^{*};\lambda_0)\right|\times I_{t,n}^{C,M,\delta}>\frac{1}{8}\epsilon\right) \tag{3.34}$$

$$+P\left(\left|\sum_{t=1}^{n}\frac{1}{2}(U_{t,n}^{+}+U_{t,n}^{-})^2e_{xx}(x_{t,n}^{*};\lambda_0)\right|\times I_{t,n}^{D,M,\delta}>\frac{1}{8}\epsilon\right), \tag{3.35}$$

where all the summands have $\lim_{M\to\infty}\lim_{\delta\to 0}\limsup_{n\to\infty}$ zero, as will be shown next.

Because

$$\begin{aligned} I_{t,n}^{A,\delta} &= I\{|U_{t,n}^{+}|<\delta\}-I\{|U_{t,n}^{+}|<\delta\}I\{|U^{-}|>\delta\} \\ &= I\{|U_{t,n}^{-}|<\delta\}-I\{|U_{t,n}^{-}|<\delta\}I\{|U^{+}|>\delta\}, \end{aligned}$$

the first summand (3.28) can be bounded above by

$$P\left(\left|\sum_{t=1}^{n}U_{t,n}^{+}e_x(\varepsilon_t;\lambda_0)\right|\times I\{|U_{t,n}^{+}|<\delta\}>\frac{1}{32}\epsilon\right) \tag{3.36}$$

$$+P\left(\left|\sum_{t=1}^{n} U_{t,n}^{+} e_x(\varepsilon_t;\lambda_0)\right| \times I\{|U_{t,n}^{+}|<\delta\} I\{|U_{t,n}^{-}|>\delta\} > \frac{1}{32}\epsilon\right) \quad (3.37)$$

$$+P\left(\left|\sum_{t=1}^{n} U_{t,n}^{-} e_x(\varepsilon_t;\lambda_0)\right| \times I\{|U_{t,n}^{-}|<\delta\} > \frac{1}{32}\epsilon\right) \quad (3.38)$$

$$+P\left(\left|\sum_{t=1}^{n} U_{t,n}^{-} e_x(\varepsilon_t;\lambda_0)\right| \times I\{|U_{t,n}^{-}|<\delta\} I\{|U_{t,n}^{+}|>\delta\} > \frac{1}{32}\epsilon\right). \quad (3.39)$$

By Markov's inequality, (3.36) is bounded by

$$\frac{32}{\epsilon}\mathrm{E}\left[\left|\sum_{t=1}^{n} U_{t,n}^{+} e_x(\varepsilon_t;\lambda_0)\right| \times I\{|U_{t,n}^{+}|<\delta\}\right]. \quad (3.40)$$

If $\mathrm{E}[e_x(\varepsilon_t;\lambda_0)] = 0$, (3.40) is zero, given that

$$\mathrm{E}\left[\left|\sum_{t=1}^{n} U_{t,n}^{+} e_x(\varepsilon_t;\lambda_0) \times I\{|U_{t,n}^{+}|<\delta\}\right|^2\right] < \infty,$$

which in turn is a consequence of independence of $U_{t,n}^{+}$ and ε_t, Minkovski inequality and Karamata's theorem,

$$\begin{aligned}
\mathrm{E}\left[\left|\sum_{t=1}^{n} U_{t,n}^{+}\right|^2 \times I\{|U_{t,n}^{+}|<\delta\}\right] \leq & n\mathrm{E}[(U_{t,n}^{+})^2 \times I\{|U_{t,n}^{+}|<\delta\}] \\
\sim & n \times (const) \times \delta^2 P(U_{t,n}^{+} > \delta) \\
= & n \times (const) \times \delta^2 P(X_{t,n}^{+} > n^{1/\alpha_0}\delta) \\
& \overset{n\to\infty}{\to} (const) \times \delta^2 < \infty.
\end{aligned}$$

For $\mathrm{E}[e_x(\varepsilon_t;\lambda_0)] \neq 0$ and $\alpha_0 < 1$, for some $C \in \mathbb{R}$, $C > \sup_x |e_x(x;\lambda_0)|$, expression (3.40) can be bounded by

$$C\frac{32}{\epsilon}\mathrm{E}\left[\left|\sum_{t=1}^{n} U_{t,n}^{+} I\{|U_{t,n}^{+}|<\delta\}\right|\right] \leq nC\frac{32}{\epsilon}\mathrm{E}[|U_{t,n}^{+}| I\{|U_{t,n}^{+}|<\delta\}]$$

$$\sim nC\frac{32}{\epsilon}\times(const)\times\delta P(|U_{t,n}^{+}|>\delta)=nC\frac{32}{\epsilon}\times(const)\times\delta P(|X_{t,n}^{+}|>n^{1/\alpha_0}\delta)$$

$$\stackrel{n\to\infty}{\to}C\frac{32}{\epsilon}\times(const)\times\delta\stackrel{\delta\to 0}{\to}0.$$

Probability in (3.37) is bounded by (3.36). Same reasoning applies to (3.38) and (3.39), where only the order of "$+$" and "$-$" signs is reverted.

Probabilities in (3.29) and (3.30) are bounded by $nP(|\varepsilon_t|>M)P(|U_{t,n}^{+}|>\delta)$ and $nP(|\varepsilon_t|>M)P(|U_{t,n}^{-}|>\delta)$, respectively. The first one has $\lim_{M\to\infty}\lim_{\delta\to 0}$ $\limsup_{n\to\infty}$ zero, as

$$\begin{aligned}nP(|\varepsilon_t|>M)P(|U_{t,n}^{+}|>\delta)&\stackrel{n\to\infty}{\sim}(const)\times P(|\varepsilon_t|>M)\\&\stackrel{M\to\infty}{\to}0.\end{aligned}$$

Same reasoning applies to (3.30) as well. Probability (3.31) can be shown to have $\lim_{M\to\infty}\lim_{\delta\to 0}$
$\limsup_{n\to\infty}$ zero by observing that it is also bounded by $nP(|\varepsilon_t|>M)P(|U_{t,n}^{+}|>\delta)$.

Probability (3.32) can be bounded by

$$\begin{aligned}&P\left(\left|\sum_{t=1}^{n}(U_{t,n}^{+}+U_{t,n}^{-})^2e_{xx}(x_t^*;\lambda_0)I\{|U_{t,n}^{+}+U_{t,n}^{-}|<2\delta\}\right|\right)\\\leq&C\frac{8}{\epsilon}\mathrm{E}\left[\left|\sum_{t=1}^{n}(U_{t,n}^{+}+U_{t,n}^{-})^2I\{|U_{t,n}^{+}+U_{t,n}^{-}|<2\delta\}\right|\right]\\\leq&C\frac{8}{\epsilon}n\mathrm{E}[(U_{t,n}^{+}+U_{t,n}^{-})^2I\{|U_{t,n}^{+}+U_{t,n}^{-}|<2\delta\}]\\\stackrel{n\to\infty}{\sim}&(const)\times C\frac{8}{\epsilon}n4\delta^2P(|U_{t,n}^{+}+U_{t,n}^{-}|>2\delta)\\\stackrel{n\to\infty}{\to}&(const)\times C\frac{8}{\epsilon}4\delta^2.\end{aligned}$$

The first row follows by the additivity of probability, the second row is due to Markov's inequality, the third row is an application of the Karamata's theorem and the fact that $0<\alpha_0<2$, and the last row is due to the fact that $U_{t,n}^{+}+U_{t,n}^{-}=n^{-1/\alpha_0}(X_{t,n}^{+}+X_{t,n}^{-})$ is an α-stable r.v.

Probabilities (3.33), (3.34) and (3.35) can be shown to have $\lim_{M\to\infty}\lim_{\delta\to 0}$

$\limsup_{n\to\infty}$ zero the same way as (3.28).

□

Proof of Lemma B7

First we show that $\{\mathcal{Q}_n^+(\cdot)\}$ is tight on $\mathbb{C}([-T,T])$. That is, for any $\epsilon > 0$, $(v', v'') \in \mathbb{R}^{P+Q}$, and $T > 0$,

$$\lim_{\epsilon\to 0^+}\limsup_{n\to\infty} P\left(\sup_{\substack{||v'-v''||<\epsilon\\ ||v'||,||v''||<T}} |\mathcal{Q}_n^+(v') - \mathcal{Q}_n^+(v'')| > \kappa\right) = 0.$$

Using mean value theorem and triangle inequality, we find that

$$\begin{aligned}|\mathcal{Q}_n^+(v') - \mathcal{Q}_n^+(v'')| &= \left|\sum_{t=1}^n e_x(\xi_{t,n};\lambda_0)n^{-1/\alpha_0}(v'-v'')^T u_{\theta,t}(\theta_0)\right| \\ &\leq \sum_{t=1}^n |[e_x(\xi_{t,n};\lambda_0) - \frac{\partial}{\partial x}e_x(\varepsilon_t;\lambda_0)]n^{-1/\alpha_0}(v'-v'')^T u_{\theta,t}(\theta_0)| \\ &\quad + \sum_{t=1}^n |e_x(\varepsilon_t;\lambda_0)n^{-1/\alpha_0}(v'-v'')^T u_{\theta,t}(\theta_0)|\end{aligned}$$

with $|\xi_{t,n} - \varepsilon_t| \leq |n^{-1/\alpha_0}(v'-v'')^T u_{\theta,t}(\theta_0)|$. Using this and the continuity of $e_x(x;\lambda_0)$, the above can be bounded by

$$\begin{aligned}&\leq K n^{-2/\alpha_0}\sum_{t=1}^n [(v'-v'')^T u_{\theta,t}(\theta_0)]^2 \\ &\quad + n^{-1/\alpha_0}\sum_{t=1}^n |e_x(\varepsilon_t;\lambda_0))(v'-v'')^T u_{\theta,t}(\theta_0)|, \qquad (3.41)\end{aligned}$$

where $K > 0$ is a constant. The result follows by showing that the sums above are $O_p(1)$. To this end, the first one can be written as

$$Kn^{-2/\alpha_0}\sum_{t=1}^n\sum_{i\neq 0} c_i^2\varepsilon_{t-i}^2 + Kn^{-2/\alpha_0}\sum_{t=1}^n\sum_{i\neq 0}\sum_{j\neq i} c_i c_j \varepsilon_{t-i}\varepsilon_{t-j}$$

$$=Kn^{-2/\alpha_0}\sum_{t=1}^{n}\sum_{i\neq 0}c_i^2\varepsilon_{t-i}^2+o_p(1),$$

where the first row follows by re-arranging the terms and the second row follows by Proposition 4.2. in Davis and Resnick (1986). Using Markov's inequality, we can evaluate the first term in the last row as

$$P\left(\left|Kn^{-2/\alpha_0}\sum_{t=1}^{n}\sum_{i\neq 0}c_i^2\varepsilon_{t-i}^2\right|>\epsilon\right)\leq \epsilon^{-\delta}K^{\delta}n^{-2\delta/\alpha_0}\mathrm{E}\left[\left|\sum_{t=1}^{n}\sum_{i\neq 0}c_i^2\varepsilon_{t-i}^2\right|^{\delta}\right]$$
$$\leq\epsilon^{-\delta}K^{-\delta}n^{-2\delta/\alpha_0}\mathrm{E}\left[\sum_{t=1}^{n}\sum_{i\neq 0}c_i^{2\delta}\varepsilon_{t-i}^{2\delta}\right]=\epsilon^{-\delta}K^{-\delta}n^{-2\delta/\alpha_0}n\mathrm{E}[\varepsilon_{t-i}^{2\delta}]\sum_{i\neq 0}c_i^{2\delta},$$

where the second inequality follows from triangle inequality. Inequality holds for any $\delta>0$ s.t. $\delta<\alpha_0/2$, thus the r.h.s. is finite and $n^{1-2\delta/\alpha_0}\to 0$ as $n\to\infty$.

For the second sum in (3.41), observe that

$$P\left(\left|n^{-1/\alpha_0}\sum_{t=1}^{n}e_x(\varepsilon_t;\lambda_0)\sum_{j\neq 0}c_j\varepsilon_{t-j}\right|>\kappa\right)$$
$$=P\left(\left|n^{-1/\alpha_0}\sum_{t=1}^{n}e_x(\varepsilon_t;\lambda_0)\sum_{j\neq 0}c_j\varepsilon_{t-j}I\left\{\left|n^{-1/\alpha_0}\sum_{j\neq 0}c_j\varepsilon_{t-j}\right|\leq\delta\right\}\right.\right.$$
$$\left.\left.+\ n^{-1/\alpha_0}\sum_{t=1}^{n}e_x(\varepsilon_t;\lambda_0)\sum_{j\neq 0}c_j\varepsilon_{t-j}I\left\{\left|n^{-1/\alpha_0}\sum_{j\neq 0}c_j\varepsilon_{t-j}\right|>\delta\right\}\right|>\kappa\right)$$
$$\leq\kappa^{-1}\mathrm{E}\left[\sum_{t=1}^{n}|e_x(\varepsilon_t;\lambda_0)|\left|n^{-1/\alpha_0}\sum_{j\neq 0}c_j\varepsilon_{t-j}\right|I\left\{n^{-1/\alpha_0}\sum_{j\neq 0}c_j\varepsilon_{t-j}\leq\delta\right\}\right]$$
$$+\kappa^{-1}\mathrm{E}\left[\sum_{t=1}^{n}|e_x(\varepsilon_t;\lambda_0)|\left|n^{-1/\alpha_0}\sum_{j\neq 0}c_j\varepsilon_{t-j}\right|I\left\{n^{-1/\alpha_0}\sum_{j\neq 0}c_j\varepsilon_{t-j}>\delta\right\}\right]$$
$$\leq\kappa^{-1}\mathrm{E}\left[\sum_{t=1}^{n}|e_x(\varepsilon_t;\lambda_0)|^{\gamma}\left|n^{-1/\alpha_0}\sum_{j\neq 0}c_j\varepsilon_{t-j}\right|^{\gamma}I\left\{n^{-1/\alpha_0}\sum_{j\neq 0}c_j\varepsilon_{t-j}\leq\delta\right\}\right]$$
$$+\kappa^{-1}\mathrm{E}\left[\sum_{t=1}^{n}|e_x(\varepsilon_t;\lambda_0)|^{\gamma}\left|n^{-1/\alpha_0}\sum_{j\neq 0}c_j\varepsilon_{t-j}\right|^{\gamma}I\left\{n^{-1/\alpha_0}\sum_{j\neq 0}c_j\varepsilon_{t-j}>\delta\right\}\right],$$

where first inequality holds by Markov's inequality, and the second inequality holds by Hölder's inequality for $\gamma = 1$ if $\alpha_0 > 1$ and $0 < \gamma < \alpha_0$ if $\alpha_0 < 1$. By the fact that $e_x(x; \lambda_0)$ is bounded and $n^{-1/\alpha_0} \sum_{j\neq 0} c_j \varepsilon_{t-j}$ is α-stable independent of $e_x(\varepsilon_t; \lambda_0)$, by Proposition A.2. (a) and (b) in Davis et al. (1992),

$$\limsup_{n\to\infty} P\left(\left| n^{-1/\alpha_0} \sum_{t=1}^{n} e_x(\varepsilon_t; \lambda_0) \sum_{j\neq 0} c_j \varepsilon_{t-j} \right| > \kappa \right)$$
$$\leq \kappa^{-1} K_1 \mathrm{E}[e_x(\varepsilon_t; \lambda_0)] \delta^{\alpha_0 - \gamma} + K_2 \delta^{-\alpha_0} P(|e_x(\varepsilon_t; \lambda_0)| > 0).$$

We have established that

$$\sup_{\substack{||v'-v''||<\epsilon \\ ||v'||,||v''||<T}} \left| \mathcal{Q}_n^+(v') - \mathcal{Q}_n^+(v'') \right| = O_p(1)$$

and it has a majorant that is linear in $(v' - v'')$, thus the result follows by letting $\epsilon \to 0$.

□

Proof of Lemma B8

After repeating Steps I and II in the proof on Lemma 1, for $\tilde{\mathcal{Q}}_n(v, w)$, this proof is essentially the same as the proof of Lemma 8 in Wu (2013). Let us define a $O(n^{1/2})$ sequence $m_n = \lfloor \sqrt{n} \rfloor$ and show that the sup of

$$\begin{aligned}
&\sum_{t=1}^{m_n-1} \sum_{t=m_n}^{n-m_n} \sum_{t=n-m_n+1}^{n} \qquad (3.42)\\
&(\ln f(\tilde{u}_t(\theta_0) + n^{-1/\alpha_0} v^T \tilde{u}_{\theta,t}(\theta_0); \lambda_0) - \ln f(\tilde{u}_t(\theta_0); \lambda_0)\\
&- \ln f(\varepsilon_t + v^T u_{\theta,t}(\theta_0); \lambda_0 - \ln f(\varepsilon_t; \lambda_0))\\
=& \sum_{t=1}^{m_n-1} \sum_{t=n-m_n+1}^{n} (\ln f(\tilde{u}_t(\theta_0) + n^{-1/\alpha_0} v^T \tilde{u}_{\theta,t}(\theta_0); \lambda_0) - \ln f(\tilde{u}_t(\theta_0); \lambda_0))\\
&+ \sum_{t=1}^{m_n-1} \sum_{t=n-m_n+1}^{n} (\ln f(\varepsilon_t + n^{-1/\alpha_0} v^T u_{\theta,t}(\theta_0); \lambda_0) - \ln f(\varepsilon_t; \lambda_0))\\
&+ \sum_{t=m_n}^{n-m_n} (\ln f(\tilde{u}_t(\theta_0) + n^{-1/\alpha_0} v^T \tilde{u}_{\theta,t}(\theta_0); \lambda_0) - \ln f(u_t(\theta_0) +
\end{aligned}$$

$$n^{-1/\alpha_0} v^T u_{\theta,t}(\theta_0); \lambda_0))$$
$$+ \sum_{t=m_n}^{n-m_n} (\ln f(\tilde{u}_t(\theta_0); \lambda_0) - \ln f(\varepsilon_t; \lambda_0))$$

is $o_p(1)$. We will show this for the first and the third rows. Same arguments can be applied to the second and the fourth as well, respectively.

The first term on the r.h.s. is bounded in absolute value by

$$\sum_{t=1}^{m_n-1} \sum_{t=n-m_n+1}^{n} |n^{-1/\alpha_0} v^T \tilde{u}_{\theta,t}(\theta_0) e_x(\tilde{u}_t(\theta_0); \lambda_0)| + \sum_{t=1}^{m_n-1} \sum_{t=n-m_n+1}^{n} |e_x'(\xi_t; \lambda_0) n^{-2/\alpha_0} (v^T \tilde{u}_{\theta,t}(\theta_0))^2| \tag{3.43}$$

In order to show that both of these terms are $o_p(1)$, we show that they converge to zero in η-mean for some $\eta > 0$, which in turn implies the convergence in probability. To this end, for $0 < \eta \leq 1$, we have

$$\begin{aligned} &\mathrm{E}\left[\left(\sum_{t=1}^{m_n-1} \sum_{t=n-m_n+1}^{n} |n^{-1/\alpha_0} v^T \tilde{u}_{\theta,t}(\theta_0) e_x(\tilde{u}_t(\theta_0); \lambda_0)|\right)^{\eta}\right] \\ \leq & n^{-\eta/\alpha_0} \sum_{t=1}^{m_n-1} \sum_{t=n-m_n+1}^{n} \mathrm{E}[|v^T \tilde{u}_{\theta,t}(\theta_0)|^{3\eta/2}]^{2/3} \mathrm{E}[|e_x(\tilde{u}_t(\theta_0); \lambda_0)|^{3\eta}]^{1/3}, \end{aligned}$$

where the Hölder's inequality has been utilized. The expectations on r.h.s. are bounded by some constant C for any $0 < \eta < 2\alpha_0/3$, so the sum is $O(m_n)$. Thus, the majorant goes to zero for any $1/2\alpha_0 \leq \eta \leq 2\alpha_0/3$.

Using similar steps, and the fact that $e_x'(\cdot; \lambda_0)$ is bounded, the η-mean of the second term in (3.43) is bounded by

$$C n^{-2\eta/\alpha_0} \sum_{t=1}^{m_n-1} \sum_{t=n-m_n+1}^{n} \mathrm{E}[|v^T \tilde{u}_{\theta,t}(\theta_0))|^{2\eta}].$$

Choosing $\alpha_0/4 \leq \eta \leq \alpha_0/2$, the η-mean has limit zero.

Combining these results, we have shown that (3.43) is $o_p(1)$.

The second row of (3.42) can be shown to be $o_p(1)$ using similar steps, and by noticing that $u_t(\theta_0)$ is independent of $u_{\theta,t}(\theta_0)$.

Next we show that the third row on the r.h.s. of (3.42) is $o_p(1)$. In absolute value, it can be bounded by

$$\begin{aligned}
&\Bigg|\sum_{t=m_n}^{n-m_n} e_x(u_t(\theta_0 + n^{-1/\alpha_0} v^T u_{\theta,t}(\theta_0)); \lambda_0) \\
&\times(\tilde{u}_t(\theta_0) - u_t(\theta_0) + n^{-1/\alpha_0} v^T(\tilde{u}_{\theta,t}(\theta_0) - u_{\theta,t}(\theta_0))) \\
&+ \frac{1}{2} e_x'(\xi_t; \lambda_0)(\tilde{u}_t(\theta_0) - u_t(\theta_0) + n^{-1/\alpha_0} v^T(\tilde{u}_{\theta,t}(\theta_0) - u_{\theta,t}(\theta_0)))^2\Bigg| \\
\leq &\sum_{t=m_n}^{n-m_n} |e_x(\varepsilon_t; \lambda_0)||\tilde{u}_t(\theta_0) - u_t(\theta_0)| \\
&+ \sum_{t=m_n}^{n-m_n} |e_x(\varepsilon_t; \lambda_0)| n^{-1/\alpha_0} |v^T(\tilde{u}_{\theta,t}(\theta_0) - u_{\theta,t}(\theta_0))| \\
&+ \sum_{t=m_n}^{n-m_n} |e_x'(\xi_t; \lambda_0)| n^{-1/\alpha_0} |v^T u_{\theta,t}(\theta_0)||\tilde{u}_{\theta,t}(\theta_0) - u_{\theta,t}(\theta_0)| \\
&+ \sum_{t=m_n}^{n-m_n} |e'(\xi_t; \lambda_0)| n^{-2/\alpha_0} |v^T u_{\theta,t}(\theta_0)||v^T(\tilde{u}_{\theta,t}(\theta_0) - u_{\theta,t}(\theta_0))| \\
&+ \sum_{t=m_n}^{n-m_n} |e_x'(\xi_t; \lambda_0)||(\tilde{u}_t(\theta_0) - u_t(\theta_0))^2 + n^{-2/\alpha_0}(v^T(\tilde{u}_{\theta,t}(\theta_0) - u_{\theta,t}(\theta_0)))^2|.
\end{aligned}$$

We show that the first and the last rows above are $o_p(1)$ by showing that for some $0 < \eta \leq 1$, their η-mean converges to zero. The rest of the rows can be shown to be $o_p(1)$ using very similar arguments.

The $\eta/2$-mean of the first row can be bounded by

$$\sum_{t=m_n}^{n-m_n} (\mathrm{E}[|e_x(\varepsilon_t; \lambda_0)|^\eta])^{1/2} (\mathrm{E}[|\tilde{u}_t(\theta_0) - u_t(\theta_0)|^\eta])^{1/2}$$

where the first expectation is a constant. For any $0 < \eta \leq \alpha_0$, the majorant has a zero limit by Lemma A2.

Using the boundedness of $e_x(\cdot)$, the last term is bounded by

$$\sum_{t=m_n}^{n-m_n} |e_x'(\xi_t; \lambda_0)||\tilde{u}_t(\theta_0) - u_t(\theta_0)|^2$$

$$
\begin{aligned}
&+ n^{-2/\alpha_0} \sum_{t=m_n}^{n-m_n} |e_x'(\xi_t;\lambda_0)||v^T(\tilde{u}_{\theta,t}(\theta_0) - u_{\theta,t}(\theta_0))|^2 \\
\leq & C \sum_{t=m_n}^{n-m_n} |\tilde{u}_t(\theta_0) - u_t(\theta_0)|^2 \\
&+ n^{-2/\alpha_0} C \sum_{t=m_n}^{n-m_n} |v^T(\tilde{u}_{\theta,t}(\theta_0) - u_{\theta,t}(\theta_0))|^2.
\end{aligned}
$$

The η-mean of the first term on the majorant side has zero limit for all $0 < \eta < \alpha_0/2$, and the same is true for the second term as well. This completes the proof that (3.42) is $o_p(1)$ for any v.

The uniform convergence on compact sets in $\mathbb{R}^{P+Q}$ follows by continuity of the functions at any v, and the boundedness of the functions.

□

4 Nonlinear predictability of asset returns[1]

4.1 Introduction

The purpose of this study is to answer a question: To what extent are U.S. stock portfolio returns and stock returns predictable? We answer this question by utilizing the properties of a noninvertible ARMA model and a simple predictability testing procedure developed in Lanne, Meitz, and Saikkonen (2013). The answer to that question is: predictability is often encountered in low frequency returns and the best one-step-ahead prediction is nonlinearly dependent on past observations.

Testing for predictability was long considered as a test for the efficient market hypothesis, as it was formalized in Fama (1970). Using this terminology, our study is related to the weak form of market efficiency. It is said to hold if future returns are not predictable using information on past prices. Lim and Brooks (2011) provides a thorough literature review of this sort of predictability. Campbell, Lo, and MacKinlay (1997), Chapter 2 reviews some of the empirical methods used to study predictability.

Since Fama (1970), theoretical dynamic asset pricing models have evolved, and predictability is no longer seen as a matter of efficiency. In fact, in consumption-based asset pricing models, predictability is an implication of agents risk aversion (Singleton, 2009, Chapter 10). These theories do not,

[1]I am grateful for the guidance and suggestions of Markku Lanne, Mika Meitz and Pentti Saikkonen. I have also benefited of the comments by Henri Nyberg, Niklas Ahlgren and Matthijs Lof. I want to thank all of the participants at the Graduate School of Finance summer school in Turku for good comments. Financial support from University of Helsinki Research funds, Yrjö Jahnsson foundation and the O-P Group research foundation is gratefully acknowledged.

however, imply autocorrelation between the returns. Predictability may be nonlinear as well.

Using ARMA models for testing for linear predictability is not meaningless, although we do not assume that it is the true data generating process. By Wold (1938), any purely nondeterministic second-order stationary processes can be presented in MA(∞) form, which in turn can be closely approximated by a finite order ARMA(P, Q) process. Many nonlinear processes admit ARMA representation. A list of some of these models can be found in Francq, Roy, and Zakoïan (2012). It is also a common practice to linearize a structural model around its steady state, and analyze the model as if it were linear. These presentations are often in the form of ARMA models. In some situations, ARMA(1,1) arises as a natural candidate for stock returns, such as in the mean reversions model of Poterba and Summers (1988) and price-trend model of Taylor (1982).

Testing for predictability with conventional causal and invertible ARMA models is, however, a non-standard procedure. Andrews and Ploberger (1994) and Nankervis and Savin (2010) provide valid tests for autocorrelation in ARMA(1,1) models under very general assumptions, but these tests are not easily extended to more general ARMA(P,Q) models. Despite being capable of producing very flexible autocorrelation structures, causal and invertible ARMA models are unable to produce any nonlinear dependence between the observations. Keeping these considerations in mind, there are a few advantages to the noninvertible ARMA model. First, on top of being as flexible with its autocorrelation structure as its invertible counterpart, the noninvertible ARMA model also imposes nonlinear structure between the observations. In a sense, the noninvertible model offers a more general model for testing the dependencies between observations. Second, autocorrelation tests can be performed with Wald or likelihood ratio test statistics that obey standard asymptotic results. Third, there is a special class of noninvertible models that produce nonlinearly predictable observations with zero autocorrelation. These so-called all-pass models are especially useful in studying stock returns, where the autocorrelation between observations is very mild at best, but it is easy to find nonlinear dependencies. For example, a mild ARCH type heroskedasticity is an example of nonlinearity, that the noninvertible ARMA model is capable of controlling.

Some theoretical justification for noninvertible ARMA model can be found in the seminal work of Hansen and Sargent (1981) and Hansen and Sargent (1991). Noninvertibility (or nonfundamentalnes) arises under assumptions that agents know more than the econometrician who tries to model the economy. More recently, similar ideas have been utilized in the asset pricing models (Kasa, Walker, and Whiteman, 2014), fiscal foresight models (Leeper, Walker, and Yang, 2013), news shock models (Blanchard and Perotti, 2002; Forni and Gambetti, 2014) and in permanent income models (Fernández-Villaverde, Rubio-Ramírez, Sargent, and Watson, 2007). What all the references above have in common is that they construct a structural economic model that has a noninvertible linearized solution. Alessi, Barigozzi, and Capasso (2011) provides a thorough survey of the literature on noninvertible economic models.

From the empirical point of view, the noninvertible model has been found to provide an adequate fit for financial time series in Andrews, Calder, and Davis (2009), Breidt, Davis, and Trindade (2001) and Huang and Pawitan (2000). The formulation and estimation theory of the noninvertible ARMA model used in this study is due to Meitz and Saikkonen (2013). As mentioned, the predictability testing procedure is introduced in Lanne et al. (2013). Model selection utilizes Ljung-Box and McLeod-Li type of portmanteau tests designed for purely noninvertible ARMA models.

Chapter 4.2 is devoted to the methodology. We introduce the noninvertible ARMA model at the depth needed to explain the predictability testing procedure. We also discuss how linear and nonlinear predictability arise from that model. The predictability tests of Lanne et al. (2013) are also introduced in Section 2, with the model selection procedure. Empirical results can be found in Chapter 4.3. Chapter 4.4 concludes.

4.2 Methodology

4.2.1 Noninvertible ARMA(P, Q) model

The formulation of the noninvertible ARMA(P, Q) model used in this study is given in Meitz and Saikkonen (2013), where its maximum likelihood (ML) estimation is also discussed. Without going into the details, we discuss the most important features of that model here.

We define noninvertible ARMA(P, Q) model as

$$a_0(B)y_t = b_0(B^{-1})\epsilon_t, \tag{4.1}$$

where $a_0(z) = 1 - a_{0,1}z - \cdots - a_{0,P}z^P$ is a P^{th} order polynomial with all the roots outside the unit circle: $a_0(z) \neq 0$ for all $|z| \leq 1$, $z \in \mathbb{C}$. Correspondingly, we define $b_0(z^{-1}) = 1 - b_0 z^{-1} - \cdots - b_Q z^{-Q}$, a Q^{th} order polynomial with all the roots outside the unit circle in complex plane, $b_0(z^{-1}) \neq 0$ for all $|z^{-1}| \leq 1$, $z \in \mathbb{C}$. B is backward shift operator, i.e. $B^k x_t = x_{t-k}$ for all $k = \ldots, -1, 0, 1, \ldots$. In contrast to the conventional causal and invertible ARMA(P, Q) model, the invertibility condition considering the MA polynomial is given in terms of argument z^{-1}. Another way of defining a noninvertible process is to have $b_0(z) \neq 0$ for all $|z| \geq 1$. ε_t is a non-Gaussian iid process, $\varepsilon_t \sim (0, \sigma^2)$. Specific high-level assumptions for this model are given and discussed in Meitz and Saikkonen (2013) and Lanne et al. (2013).

4.2.2 Linear and nonlinear predictability

Conditional expectation is denoted by $\mathrm{E}_t[X_{t+1}] \stackrel{def.}{=} \mathrm{E}[X_{t+1}|X_{t-s}, s \geq 0]$. Predictability of the process y_t means that its conditional mean is not constant, $E_{t-1}[y_t] \neq 0$, for some $t = \ldots, -1, 0, 1, \ldots$.

It is well known that, for causal and invertible ARMA models, lack of autocorrelation implies the unpredictability of the process. This is not the case with the noninvertible ARMA model. This can be seen by looking at the spectral density function of process y_t in (4.1):

$$f_y(\zeta) = \frac{\sigma_0^2}{2\pi} \frac{b_0(e^{i\zeta}) b_0(e^{-i\zeta})}{a_0(e^{-i\zeta}) a_0(e^{i\zeta})}.$$

As pointed out in Lanne et al. (2013), an interesting special case is one where $P = Q$ and $a_0(z) = b_0(z)$. In this case the spectral density is not dependent on ζ, and the autocorrelation is zero. This is the so-called all-pass model, studied for example by Breidt et al. (2001) and Andrews, Davis, and Breidt (2006). It is important to notice that, even for the all-pass model, the observations are dependent. In the fashion of (4.1), the model can be written as

$$a_0(B) y_t = a_0(B^{-1}) \epsilon_t,$$

where the polynomials do not cancel out since the polynomial on the right hand side has an argument z^{-1} instead of z. Formal justification for the nonlinear predictability of the all-pass model can be shown in the same way as in Lanne et al. (2013), Appendix A, where it has been shown for noninvertible ARMA(1,1) model. The predictability of y_t boils down to the question, if $a_0(z) = b_0(z) = 1$, and, $y_t = \epsilon_t$.

4.2.3 Predictability testing procedure

The noninvertible ARMA model has a major advantage in hypothesis testing over its invertible counterpart. Consider a causal and invertible ARMA(P, P) model

$$a_0(B)y_t = b_0(B)\varepsilon_t, \tag{4.2}$$

that satisfies the assumptions listed below (4.1). Testing for linear predictability would result in testing a hypothesis $a_0(z) = b_0(z)$ in (4.2). Note that under the null, the AR and MA polynomials would cancel out and there are several parameters that would be present only under the alternative hypothesis. Such a setting results in unconventional testing theory which has been studied by Andrews and Ploberger (1994) and Nankervis and Savin (2010), and is very hard to generalize to models larger than ARMA(1,1). The noninvertible ARMA model allows us to test for linear predictability in a standard manner because AR and MA parameters are present also under the null.

ML estimation theory for the model (4.1), with an ARCH(R) type of error terms, is derived in Meitz and Saikkonen (2013), which includes a set of high-level assumptions on (4.1). These assumptions are very standard in nature (see e.g. Lii and Rosenblatt (1996), Breidt et al. (2001) or Andrews et al. (2009).) The set of assumption in Meitz and Saikkonen (2013) is tailored to the formulation of the noninvertible model (4.1), which differs slightly from the formulation in other references. Also, the assumptions used here account for the ARCH(R) type error term, which is not present in the other references mentioned above, nor in our model. We use ML estimation method with re-scaled Student's t-distribution, which is well in line with the assumptions.

Next, we set some notation. Let $\tilde{\theta} = (\tilde{a}, \tilde{b})$ with $\tilde{a} = (\tilde{a}_1, \ldots, \tilde{a}_P)$ and $\tilde{b} = (\tilde{b}_1, \ldots, \tilde{b}_Q)$ be the ML estimators of the AR and MA parameters of (4.1). Let $\tilde{\delta} = (\tilde{\theta}, \tilde{\sigma}, \tilde{\lambda})$ denote the $(P + Q + 1 + d \times 1)$ parameter vector of all the parameters of the model, where $\tilde{\sigma}$ is the standard deviation of the error process ε_t. Density function of the iid error term ε_t is $\sigma_0^{-1} f_\varepsilon(\sigma_0^{-1}x; \lambda_0)$, is (possibly) dependent on λ_0. In our case, λ_0 is the degrees of freedom parameter of Student's t-distribution.

Local maximizer of approximative likelihood function[2]

$$L(\delta) = n^{-1} \sum_{t=1}^{n} (\log f_\varepsilon(\sigma^{-1}\varepsilon_t; \lambda) - \frac{1}{2} \log \sigma^2)$$

[2]The details for this approximate likelihood functions are in Meitz and Saikkonen (2013)

is consistent and asymptotically normally distributed,

$$n^{1/2}(\tilde{\delta} - \delta_0) \xrightarrow{d} N(0, \mathcal{I}(\delta_0)^{-1}),$$

as the sample size n goes to infinity. The positive definite covariance matrix $\mathcal{I}(\delta_0)^{-1}$ is obtained as a probability limit of the Hessian of the log-likelihood function in the usual manner:

$$-\frac{\partial^2}{\partial \delta \partial \delta^T} L(\tilde{\delta}) \stackrel{def}{=} \tilde{\mathcal{J}} \xrightarrow{p} \mathcal{I}(\delta_0). \tag{4.3}$$

In the fashion of Lanne et al. (2013) and in the light of the previous subsection, we have two hypotheses we wish to test. The first is the test for linear predictability. We test for the all-pass hypothesis against an unrestricted non-invertible ARMA(P, P) model,

$$H_{ap}: \; a_{0,p} = b_{0,p} \quad \text{for all } p = 1, \ldots, P, \text{ in (4.1).}$$

Testing H_{ap} involves jointly testing P linear restrictions and a corresponding Wald test is

$$W_{ap} = (\tilde{a} - \tilde{b})^T (n^{-1} R (\mathcal{J}_{11} - \mathcal{J}_{12} \mathcal{J}_{22}^{-1} \mathcal{J}_{12})^{-1} R^T)^{-1} (\tilde{a} - \tilde{b}) \xrightarrow{d} \chi^2_P, \quad \text{under } H_{ap}$$

with

$$R = \begin{pmatrix} I_{P \times P} & -I_{P \times P} \end{pmatrix},$$

$I_{P \times P}$ is a $(P \times P)$ identity matrix, and $\mathcal{I}_{ij}$ are the blocks of

$$\mathcal{I}(\tilde{\delta}) = \begin{pmatrix} \mathcal{I}_{11} & \mathcal{I}_{12} \\ \mathcal{I}_{21} & \mathcal{I}_{22} \end{pmatrix}.$$

If $\tilde{\delta}_{ap} = (\tilde{\theta}_{ap}, \tilde{\sigma}_{ap}, \tilde{\lambda}_{ap})$ is the vector of ML estimates of the constrained model satisfying H_{ap}, then the likelihood ratio test for H_{ap} is

$$LR_{ap} = 2n[L(\tilde{\delta}) - L(\tilde{\delta}_{ap})] \xrightarrow{d} \chi^2_P \quad \text{under } H_{ap}.$$

The second hypothesis we wish to test is the unpredictability hypothesis

$$H_{iid}: \ a_{0,p} = b_{0,p} = 0, \quad \text{for all } p = 1, \ldots, P, \text{ in (4.1)},$$

against the alternative of H_{ap}. After noninvertible ARMA(P, P) model in (4.1) has been estimated under the H_{ap}, Wald test reads as

$$W_{iid} = n\tilde{\theta}_{ap}^T (\tilde{\mathcal{J}}_{ap,11} - \tilde{\mathcal{J}}_{ap,12}\tilde{\mathcal{J}}_{ap,2}^{-1}\tilde{\mathcal{J}}_{ap,12})\tilde{\theta}_{ap} \xrightarrow{d} \chi_P^2, \quad \text{under } H_{iid},$$

where $\tilde{\mathcal{J}}_{ap,11} - \tilde{\mathcal{J}}_{ap,12}\tilde{\mathcal{J}}_{ap,2}^{-1}\tilde{\mathcal{J}}_{ap,12}$ in the middle is the upper-left block of the inverse of the Hessian matrix of the model estimated under H_{ap}.

Likelihood ratio test is formulated in the usual manner,

$$2n[L(\tilde{\delta}_{ap}) - L(\tilde{\delta}_{iid})] \xrightarrow{d} \chi_P^2, \quad \text{under } H_{iid},$$

where $\tilde{\delta}_{iid}$ is a vector of estimates under H_{iid}, $\tilde{\delta}_{iid} = (\mathbf{0}_{(P\times 1)}^T, \sigma_{iid}^T, \lambda_{iid})$. The drawback of using the likelihood ratio statistics is, as usual, that the model must be estimated under the null hypothesis, in addition to the alternative.

4.2.4 Model selection

Model selection can be based on a Box and Pierce (1970) and McLeod and Li (1983) type of portmanteau tests Q_{ac} and Q_{hs}. In the case of noninvertible ARMA models, the properties of these tests have been derived in Nyholm (2017). In order to execute the tests, we have to calculate the residuals of the fitted models. By the way the model (4.1) is written, the residuals must be solved recursively top-down. Assume that we observed $\{y_t\}_{t=1-P}^n$, and let u_t denote the residual that satisfies $y_t - \tilde{a}_1 y_{t-1} - \cdots - \tilde{a}_P y_{t-P} = u_t - \tilde{b}_1 u_{t+1} - \cdots - \tilde{b}_P u_{t+P}$ for $t = n, \ldots, 1$, with initialization $u_{n+1} = \cdots = u_{n+P} = 0$.

In order to test for the autocorrelation, we use test statistic

$$Q_{ac} = n\rho_{ac}^T \tilde{\Omega}_{ac}^{-1} \rho_{ac},$$

where ρ_{ac} is a $(m \times 1)$ vector of empirical autocorrelation coefficients of u_t,

$$\rho_{i,ac} = \frac{\sum_{t=i+1}^n u_t u_{t-i}}{\sum_{t=1}^n u_t u_t}, \quad \text{and} \quad \rho_{ac} = (\rho_{1,ac}, \ldots, \rho_{m,ac})$$

for some m. The covariance matrix in the middle is

$$\tilde{\Omega}_{ac} = I_{m\times m} - \tilde{\Lambda}_m \tilde{\mathcal{J}}^{11} \tilde{\Lambda}_m^T,$$

and $\tilde{\mathcal{J}}$ is the upper diagonal block of $\tilde{\mathcal{J}}^{-1}$, given in (4.3), and

$$\tilde{\Lambda}_m = \begin{pmatrix} \tilde{\psi}_0^{(a)} & \dots & \tilde{\psi}_{1-P}^{(a)} & -\tilde{\psi}_0^{(b)} & \dots & -\tilde{\psi}_{1-Q}^{(b)} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \tilde{\psi}_{m-1}^{(a)} & \dots & \tilde{\psi}_{m-P}^{(a)} & -\tilde{\psi}_{m-1}^{(b)} & \dots & -\tilde{\psi}_{m-Q}^{(b)} \end{pmatrix},$$

where $\tilde{\psi}_i^{(a)}$ and $\tilde{\psi}_i^{(b)}$ are the coefficients of the power series expansions of

$$\frac{1}{1-\tilde{a}_1 z - \cdots - \tilde{a}_P z^P} = \sum_{j=0}^{\infty} \tilde{\psi}_j^{(a)} z^j \quad \text{and} \quad \frac{1}{1-\tilde{b}_1 z - \cdots - \tilde{b}_P z^P} = \sum_{j=0}^{\infty} \tilde{\psi}_j^{(b)} z^j,$$

respectively. For negative indices i, $\psi_i^{(\cdot)} = 0$. If we conduct the test for noninvertible ARMA(1,1) model, we use

$$\Lambda_m^{(1,1)} = \begin{pmatrix} 1 & -1 \\ \tilde{a}_1 & -\tilde{b}_1 \\ \vdots & \vdots \\ \tilde{a}_1^{m-1} & -\tilde{b}_1^{m-1} \end{pmatrix}.$$

Under the null hypothesis of correctly specified ARMA model in (4.1), then asymptotically $Q_{ac} \sim \chi^2_m$, and we reject the null with confidence level α if $Q_{ac} > k_{1-\alpha}$, with $P(\chi^2_m \geq k_{1-\alpha}) = \alpha$.

Heteroskedasticity is tested by using

$$Q_{hs} = n\rho_{hs}^T \rho_{hs},$$

where ρ_{hs} is a $(m \times m)$ vector of empirical autocorrelations of u_t^2,

$$\rho_{i,hs} = \frac{\sum_{t=i+1}^{n} (u_{t-i}^2 - \tilde{\sigma}^2)(u_t^2 - \tilde{\sigma}^2)}{\sum_{t=1}^{n} (u_t^2 - \tilde{\sigma}^2)} \quad \text{and} \quad \rho_{hs} = (\rho_{1,hs}, \dots, \rho_{m,hs})$$

for some m. If the model is correctly specified, $Q_{hs} \sim \chi^2_m$ asymptotically, and we can calculate the critical values for the tests the same way as above.

4.3 Predictability of the asset return portfolios

As pointed out by Lanne et al. (2013), noninvertible ARMA model has several advantages against other methods of predictability testing. Linear predictability can be tested in a straightforward manner by calculating the sample autocorrelation function and comparing it to a suitable critical value. Usually this value is the asymptotic 95% critical value calculated under the assumption of iid returns. This method does, however, suffer from poor power properties, as explained in Campbell et al. (1997), Chapter 2.

Imposing more structure on the testing procedure improves the power of the testing. A natural way of doing so is to impose ARMA(1,1) structure to the asset returns, which arises naturally from the price-trend model of Taylor (1982) and mean reversion model of Poterba and Summers (1988). As pointed out earlier, it is a nonstandard procedure to test for the autocorrelation in the ARMA(1,1) model, as the AR and MA parameters are present only under the hypothesis of linear predictability.

Nonlinear predictability is searched using some particular nonlinear model. For example, sign predictability or smooth transition models have been popular in recent literature. Also GARCH-in-mean models have been used to study the relation between volatility and excess returns, see e.g. Bollerslev, Engle, and Wooldridge (1988). Another branch of literature derives nonlinear predictability in a closed form from a theoretical asset pricing model. A classical example of this is the estimation of the parameters of the stochastic Euler equation, for example in Hansen and Singleton (1982). Our method, falls in the first category, as we do not test for a hypothesis implied by a structural economic model, although the noninvertible ARMA model may be compatible with such a model.

It is superior to the conventional causal and invertible ARMA model because it incorporates all the same autocorrelation structures, but the test for linear predictability is a standard procedure. Furthermore, if no linear predictability is found, we may still conduct a test for nonlinear predictability. For conventional ARMA models, the lack of linear predictability always implies the lack of any predictability.

The main aim of this study is to test whether or not there is any predictability in U.S. stock returns that can be detected with a noninvertible ARMA model, and to what extent.

4.3.1 U.S. stock portfolios returns

We start our analyses by evaluating the fit of the noninvertible ARMA model to the return series of 10 value-weighted portfolios formed on size. Return series are continuously compounded using monthly data from Kenneth French's data library, which combines data from the NYSE, AMEX and NASDAQ. The data spans from the first quarter of 1947 to the second quarter on 2019. We present estimation results of the noninvertible ARMA(1,1) model for the market portfolio and 10 value-weighted size-ordered portfolios, ranked decile-by-decile, see Table 4.1. The estimated AR and MA coefficients are very close to each other for all the portfolios, which points towards the all-pass hypothesis. Parameters are also relatively far from zero, compared to their estimated standard deviations. The residuals of these portfolios do not depict any signs of autocorrelation, which can be seen by calculating the Q_{ac} statistic for testing the autocorrelation. The p-values of these tests are so large that the hypothesis of no autocorrelation cannot be rejected on any meaningful significance level. There are some signs of heteroskedasticity in the residuals of two portfolios. These are the portfolios of the smallest companies and the group in the second decile. This suggests that a larger noninvertible ARMA model might be more adequate for these portfolios, and indeed, for the Decile 2 portfolio, a noninvertible ARMA(2,2) model is capable of capturing the dependencies in the residuals. The estimated model for this portfolio is $(1 - 0.621B + 0.649B^2)y_t = (1 - 0.644B^{-1} + 0.604B^{-2})\epsilon_t$. Now the p-values for the autocorrelation in the residuals and in the squared residuals for $m = 12$ are 0.758 and 0.638, respectively, so there is no sign of dependencies left in the residuals. Numbers are similar for other lag lengths as well.

Next we concentrate on the quarterly returns of 46 different stock return portfolios. Quarterly returns are again composed from monthly data, spanning from January 1947 to December 2017. There are 284 observations in each portfolio. We use 25 portfolios formed according to the size and book-to-market values of the companies in the stock exchanges (5×5). Market portfolio represents the return on the value weighted market portfolio of all of the companies in all of the stock exchanges, in excess to risk-free rate. We use 5 industry portfolios, where the companies are divided into Consumer, Manufacturing, High Tech, Health Care or other industries. We also analyze 10 industry portfolios where the companies are classified as No Durable, Durable, Manufacturing, Energy, High Tech, Telecommunications, Health Care, Shop-

					$Q_{ac,T}$			$Q_{hs,T}$		
Portfolio	a	b	σ^2	λ	5	9	12	5	9	12
Market	.775	.758	8.074	5.012	.807	.625	.746	.812	.845	.679
	(.037)	(.039)	(.436)	(0.793)						
Decile 1	−0.489	−0.542	12.448	6.750	.551	.491	.464	.121	.164	.001
	(.151)	(.138)	(.890)	(2.271)						
Decile 2	−0.574	−0.576	11.782	6.415	.792	.746	.740	.345	.330	.012
	(.045)	(.032)	(.594)	(1.929)						
Decile 3	−0.418	−0.419	11.122	6.382	.568	.551	.581	.268	.494	.280
	(.055)	(.055)	(.580)	(2.135)						
Decile 4	.746	.721	7.679	5.152	.974	.963	.963	.387	.500	.856
	(.050)	(.035)	(.573)	(1.181)						
Decile 5	.742	.821	10.816	4.958	.636	.615	.743	.290	.415	.209
	(.060)	(.061)	(.541)	(1.621)						
Decile 6	−0.462	−0.484	10.392	4.667	.853	.753	.846	.723	.839	.802
	(.038)	(.050)	(.525)	(1.150)						
Decile 7	.723	.806	9.795	4.953	.719	.571	.672	.221	.288	.308
	(.050)	(.055)	(.549)	(0.634)						
Decile 8	.691	.806	10.014	3.999	.794	.665	.775	.347	.490	.455
	(.046)	(.043)	(.466)	(0.819)						
Decile 9	.669	.692	8.671	4.039	0.466	.391	.942	.580	.353	.411
	(.049)	(.047)	(.488)	(.666)						
Decile 10	0.789	0.734	7.678	4.844	1.000	.783	.835	.754	.941	.900
	(.033)	(.047)	(.409)	(1.031)						

Table 4.1: The noninvertible ARMA(1,1) model has been estimated to 11 stock return index series: the CRSP market portfolio returns and returns of 10 value weighted-portfolios formed on size. Table indicates the parameter estimates and their standard errors. Test statistics $Q_{ac,T}$ and $Q_{hs,T}$ have been calculated from the residuals of the fitted models and their p-values have been reported for three different lag lengths m for each test.

ping, Utilities and other industries. On top of these, we analyze four Dow-Jones Average indices: Industrial, Composite, Transportation and Utility, and the S&P 500 Index.

Linear and nonlinear predictability has been tested in all the 46 portfolios using a noninvertible ARMA(1,1) model. The results are listed in Tables 4.2 and 4.3 in the Appendix for the Likelihood ratio test and Wald test, respectively. Of all the 46 portfolios, the all-pass hypothesis is rejected in seven cases using the likelihood ratio test and ten cases using the Wald test, with 5% significance level. All of these rejections suggest that there might be linear predictability in these portfolios. Almost all of these portfolios belong to the set of 25 portfolios formed according to size and book-to-market values. Eyeballing, the test results do not resemble any clear systematic relation, but linear predictability is never found for the portfolios of the high book-to-market firms. Of those portfolios, where the likelihood ratio test did not reject the all-pass hypothesis, the iid hypothesis was rejected 14 times. These 14 portfolios exhibit signs of nonlinear predictability. For example, the CRSP market portfolio belongs to this set. Using the Wald test, this set is even larger. Altogether, 28 portfolios show sign of nonlinear predictability, but no sign of linear predictability. For example, all except one of the CRSP 10 industry portfolios are nonlinearly predictable, but not linearly, as well as three out of four Dow-Jones Average indices and the S&P500 index.

4.3.2 S&P 500 financial sector firms' stock returns

We have analyzed the stock returns of 21 financial sector companies in the S&P 500 list. Returns are continuously compounded from monthly stock prices, adjusted for dividend payments. Data availability restricts the set of companies we can analyze. We only analyze companies with price data available prior to 1990 and spans until the end of 2017. Names of the companies are listed in Table 4.4 in Appendix, together with the test results.

We start by estimating noninvertible ARMA(1,1) model for the stock returns of each of the companies. This model is extended to nonivertible ARMA(2,2) model, if either Q_{ac} or Q_{hs} test rejects at 95% significance level. P-values of the Wald test for linear and nonlinear predictability are reported for all the estimated models in Table 4.4. Among the returns on these companies' stocks, we find very little signs of autocorrelation. Noninvertible ARMA(1,1) model seems adequate for 14 out of 21 occasions. H_{ap} is not rejected in any of these cases. H_{iid} is rejected in 9 cases out of these 14.

If the noninvertible ARMA(1,1) is not capable of controlling for the heteroskedasticity in the data, the noninvertible ARMA(2,2) model is rarely better. For Comerica Inc., extending noninvertible ARMA(1,1) to ARMA(2,2) helps to control for the dependencies in the residuals. In this case, H_{ap} and H_{iid} are clearly rejected.

4.4 Conclusions

We have applied the two-stage predictability testing procedure of Lanne et al. (2013) to a set of returns on stock portfolios and financial sector firms stocks. The results show clear signs of predictability in asset returns, which is not linear in many cases. This result is well in line with the dynamic consumption-based asset pricing literature, see e.g. Singleton (2009), Chapter 9. The second contribution is a straightforward extension of the testing procedure of Lanne et al. (2013) to incorporate larger models than nonivertible ARMA(1,1).

We used diagnostic tests for the residuals of noninvertible ARMA(P, Q) model to analyze the goodness of fit of the model. Q_{ac} and Q_{hs} statistics for testing autocorrelation in the residuals and squared residuals give very valuable insights into the goodness of fit of the noninvertible ARMA model. We showed how these tests works in practice in a small-scale example, and results were encouraging. Noninvertible ARMA models were successful in controlling the dependence structure in the data according to these tests, many of the return series.

Nonlinear predictability was not exceptional among the data we used. Our testing procedure revealed linear predictability in a few cases, but around half of the portfolios showed signs of some kind of predictability. Among different portfolio classes, we were not able to see any systemic pattern, although linear predictability seems to be less likely to occur among the high book-to-market value firm portfolios, or portfolios formed by industries. Also, large financial sector firms showed very little sign of autocorrelation, but nonlinear predictability was present in most of the cases in which noninvertible ARMA(1,1) model was shown to be an adequate model.

References

Alessi, L., M. Barigozzi, and M. Capasso (2011). Non-fundamentalness in structural econometric models: A review. *International Statistical Review 79*, 16–47.

Andrews, B., M. Calder, and R. A. Davis (2009). Maximum likelihood estimation for α-stable autoregressive processes. *Annals of Statistics 37*, 1946–1982.

Andrews, B., R. A. Davis, and F. J. Breidt (2006). Maximum likelihood estimation for all-pass time series models. *Journal of Multivariate Analysis 97*, 1638 – 1659.

Andrews, D. W. and W. Ploberger (1994). Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica 62*, 1383–1414.

Blanchard, O. and R. Perotti (2002). An empirical characterization of the dynamic effects of changes in government spending and taxes on output. *Quarterly Journal of Economics 117*, 1329–1368.

Bollerslev, T., R. F. Engle, and J. M. Wooldridge (1988). A capital asset pricing model with time-varying covariances. *Journal of Political Economy 96*, 116–131.

Box, G. E. and D. A. Pierce (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association 65*, 1509–1526.

Breidt, F. J., R. A. Davis, and A. A. Trindade (2001). Least absolute deviation estimation for all-pass time series models. *Annals of Statistics 29*, 919–946.

Campbell, J. Y., A. W.-C. Lo, and A. C. MacKinlay (1997). *The econometrics of financial markets*, Volume 2. Princeton University Press, Princeton, NJ.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The journal of Finance 25*, 383–417.

Fernández-Villaverde, J., J. F. Rubio-Ramírez, T. J. Sargent, and M. W. Watson (2007). ABCs (and Ds) of understanding VARs. *American Economic Review 97*, 1021–1026.

Forni, M. and L. Gambetti (2014). Sufficient information in structural VARs. *Journal of Monetary Economics 66*, 124–136.

Francq, C., R. Roy, and J.-M. Zakoïan (2012). Diagnostic checking in arma models with uncorrelated errors. *Journal of the American Statistical Association 100*, 532–544.

Hansen, L. P. and T. J. Sargent (1981). Exact linear rational expectations models: Specification and estimation. Technical report, Federal Reserve Bank of Minneapolis.

Hansen, L. P. and T. J. Sargent (1991). Two difficulties in interpreting vector autoregressions. In L. P. Hansen, T. J. Sargent, J. Heaton, A. Marcet, and W. Roberds (Eds.), *Rational expectations econometrics*. Oxford: Westview Press Boulder, CO.

Hansen, L. P. and K. J. Singleton (1982). Generalized instrumental variables estimation of nonlinear rational expectations models. *Econometrica 50*, 1269–1286.

Huang, J. and Y. Pawitan (2000). Quasi-likelihood estimation of non-invertible moving average processes. *Scandinavian Journal of Statistics 27*, 689–702.

Kasa, K., T. B. Walker, and C. H. Whiteman (2014). Heterogeneous beliefs and tests of present value models. *Review of Economic Studies 81*, 1137–1163.

Lanne, M., M. Meitz, and P. Saikkonen (2013). Testing for linear and nonlinear predictability of stock returns. *Journal of Financial Econometrics 11*, 682–705.

Leeper, E. M., T. B. Walker, and S.-C. S. Yang (2013). Fiscal foresight and information flows. *Econometrica 81*, 1115–1145.

Lii, K.-S. and M. Rosenblatt (1996). Maximum likelihood estimation for non-gaussian nonminimum phase ARMA sequences. *Statistica Sinica 6*, 1–22.

Lim, K.-P. and R. Brooks (2011). The evolution of stock market efficiency over time: a survey of the empirical literature. *Journal of Economic Surveys 25*, 69–108.

McLeod, A. I. and W. K. Li (1983). Diagnostic checking ARMA time series models using squared-residual autocorrelations. *Journal of Time Series Analysis 4*, 269–273.

Meitz, M. and P. Saikkonen (2013). Maximum likelihood estimation of a non-invertible ARMA model with autoregressive conditional heteroskedasticity. *Journal of Multivariate Analysis 114*, 227–255.

Nankervis, J. C. and N. E. Savin (2010). Testing for serial correlation: generalized andrews–ploberger tests. *Journal of Business & Economic Statistics 28*, 246–255.

Nyholm, J. (2017). Residual-based diagnostic tests for noninvertible ARMA models. Working Papers 416, Helsinki Center of Economic Research discussion papers.

Poterba, J. M. and L. H. Summers (1988). Mean reversion in stock prices. *Journal of Financial Economics 22*, 27 – 59.

Singleton, K. J. (2009). *Empirical dynamic asset pricing: model specification and econometric assessment*. Princeton University Press.

Taylor, S. J. (1982). Tests of the random walk hypothesis against a price-trend hypothesis. *Journal of Financial and Quantitative Analysis 17*, 37–61.

Wold, H. (1938). *A Study in the Analysis of Stationary Time Series*. Almqvist & Wiksells boktrycheri-a.-b.

Table appendix

4.A Tables

LR test		P-values			P-values			P-values	
		H_{ap}	H_{iid}		H_{ap}	H_{iid}		H_{ap}	H_{iid}
CRSP Market Portfolio		0,87	0,01						
CRSP 25 portfolios formed				CRSP 5 industry			Dow-Jones		
on size and book to market				portfolios			Averages		
Size	B / M			Consumer	0,89	0,10	Industrial	0,66	0,09
Small	Low	0,03	1,00	Manuf	0,96	0,01	Composite	0,96	0,01
	2	0,31	0,26	HiTec	1,00	0,64	Transportation	0,07	0,05
	3	1,00	0,11	Hlth	0,97	1,00	Utility	0,04	0,02
	4	0,17	1,00	Other	0,32	0,04			
	High	0,06	1,00						
2	Low	0,04	0,03	CRSP 10 Industry					
	2	0,03	1,00	Portfolios			SP 500 Index	0,58	0,14
	3	0,04	0,16	noDur	0,80	0,00			
	4	0,19	0,43	Durbl	0,33	0,10			
	High	0,06	1,00	Manuf	0,04	0,02			
3	Low	0,06	0,04	Energy	0,55	0,01			
	2	1,00	0,01	HiTec	0,36	0,18			
	3	0,10	0,04	Telcm	0,12	0,00			
	4	0,09	0,30	Shop	0,04	0,33			
	High	0,06	1,00	Hlth	0,83	0,61			
4	Low	0,14	0,00	Utils	0,46	0,08			
	2	0,23	0,01	Other	0,14	0,04			
	3	0,24	0,19						
	4	0,01	0,03						
	High	0,12	0,01						
Large	Low	0,38	0,13						
	2	0,13	0,07						
	3	0,05	0,08						
	4	0,23	0,48						
	High	0,85	0,30						

Table 4.2: Likelihood ratio tests for linear and nonlinear predictability. We have tested for linear and nonlinear predictability using the LR-test and the noninvertible ARMA(1,1) model. The data is quarterly measured, and it spans the years from 1947 to 2017. Linear predictability can be found in seven portfolios using the 5% significance level. Nonlinear predictability was found in 14 portfolios.

Wald test		P-values			P-values			P-values	
		H_{ap}	H_{iid}		H_{ap}	H_{iid}		H_{ap}	H_{iid}
CRSP Market Portfolio		1,00	0,00						
CRSP 25 portfolios Formed				CRSP 5 industry			Dow-Jones		
on Size and Book to Market				Portfolios			Averages		
Size	B / M			Consumer	0,87	0,00	Industrial	0,96	0,00
Small	Low	0,01	0,00	Manuf	1,00	0,00	Composite	0,96	0,00
	2	0,03	0,14	HiTec	0,95	0,67	Transportation	0,11	1,00
	3	0,20	1,00	Hlth	0,98	0,00	Utility	0,05	0,00
	4	0,11	0,00	Other	0,84	1,00			
	High	0,05	0,00						
2	Low	0,87	0,00	CRSP 10 Industry					
	2	0,03	0,00	Portfolios			SP 500 Index	0,55	0,00
	3	0,01	0,00	noDur	0,89	0,00			
	4	0,13	0,00	Durbl	0,56	0,00			
	High	0,02	0,00	Manuf	0,04	0,00			
3	Low	0,05	0,00	Energy	0,98	0,00			
	2	0,24	1,00	HiTec	0,94	0,00			
	3	0,08	0,00	Telcm	0,10	0,00			
	4	0,02	0,00	Shop	0,61	0,00			
	High	0,04	0,00	Hlth	0,98	0,00			
4	Low	0,11	0,00	Utils	0,45	0,01			
	2	0,21	0,00	Other	0,84	1,00			
	3	0,04	0,99						
	4	0,00	0,00						
	High	0,09	0,00						
Large	Low	0,40	0,00						
	2	0,62	0,97						
	3	0,53	0,00						
	4	0,49	0,76						
	High	0,86	0,03						

Table 4.3: Wald tests for linear and nonlinear predictability. We have executed Wald tests for testing the hypothesis on linear and nonlinear predictability using the ARMA(1,1) model. The p-values are reported. The data is quarterly measured, and it spans the years from 1947 to 2017. Linear predictability can be found in seven portfolios using the 5% significance level. Nonlinear predictability was found in 28 portfolios.

	Noninvertible ARMA(1,1)				Noninvertible ARMA(2,2)			
	P-values				P-values			
	H_{ap}	H_{iid}	Q_{ac}	Q_{hs}	H_{ap}	H_{iid}	Q_{ac}	Q_{hs}
S&P 500 Financial Companies								
Aflac Inc.	0.593	$1.59e^{-4}$	0.6478	0.969				
American Express Co.	0.58	$8.46e^{-14}$	0.9563	0.025				
American International Group	0.228	$< 1e^{-20}$	0.0952	$8.48e^{-4}$				
AON plc.	0.865	0.017	0.3883	0.055				
Arthur J. Gallagher & Co.	0.545	0.074	0.6047	0.882				
BB&T Corp.	0.357	$5.70e^{-11}$	0.8853	$5.30e^{-7}$	0.179	$< 1e^{-20}$	0.585	0.016
Charles Schwab Corp.	0.199	0.55	0.2786	0.004	$1.64e^{-5}$	$< 1e^{-20}$	0.462	$7.38e^{-7}$
Chubb Limited	0.900	0.757	0.1452	0.784				
Cincinnati Financial	0.876	0.029	0.9014	0.687				
Comerica Inc.	0.668	0.229	0.7043	0.004	$9.78e^{-4}$	$< 1e^{-20}$	0.762	0.393
Fifth Third Bancorp	0.838	$4.53e^{-11}$	0.0932	0.004	$1.41e^{-4}$	$< 1e^{-20}$	0.455	$4.56e^{-5}$
Franklin Resources Inc.	0.911	0.704	0.9807	0.902				
Huntington Bancshares	0.743	0.215	0.7342	$4.60e^{-10}$	$2.22e^{-16}$	$< 1e^{-20}$	0.013	0.018
Lincoln National Corp.	0.455	$< 1e^{-20}$	0.5789	0.281				
Marsh & McLennan Comapnies Inc.	0.960	0.277	0.3391	0.817				
Northern Trust Corp.	0.960	$2.22e^{-15}$	0.7175	0.953				
Peoples United Financial Inc.	0.070	$< 1e^{-20}$	0.218	0.002				
Progressive Corp.	1.000	0.004	0.3397	0.041	0.001	$< 1e^{-20}$	0.436	0.035
Raymond James Financial Inc.	0.482	0.877	0.9814	0.805				
S&P Global Inc.	0.116	$3.93e^{-7}$	0.4149	0.018	0.801	$< 1e^{-20}$	0.579	0.011
State Street Corp.	0.635	$1.30e^{-14}$	0.1717	0.064				

Table 4.4: Wald test for linear and nonlinear predictability of S&P 500 Financial sector stock returns.We have analyzed 21 S&P500 list financial companies' quarterly stock return series. Returns are adjusted for dividend payments, and the companies are chosen such that there was price data available prior to 1990 to the end of 2017. We have tested for linear and nonlinear predictability using the ARMA(1,1) model whenever the Q_{ac} and Q_{hs} tests do not reject. When there is a rejection, we have expanded to the noninvertible ARMA(2,2) model. The p-values of the Wald tests are reported. We found ARMA(1,1) as a suitable model for the testing purpose for 14 return series, and in nine of these cases there were signs of nonlinear predictability.